

Open Research Online

The Open University's repository of research publications and other research outputs

The Dialogic Nature Of Online Discourse: A Corpus Analysis Of Online Discussions

Thesis

How to cite:

Chua, Shi Min (2021). The Dialogic Nature Of Online Discourse: A Corpus Analysis Of Online Discussions. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 2020 Shi Min Chua



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Version of Record

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.21954/ou.ro.000124f8>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

The dialogic nature of online discourse: A corpus analysis of online discussions

Shi Min Chua

蔡思敏

A thesis submitted to
The Open University
for the degree of
DOCTOR OF PHILOSOPHY

Faculty of Wellbeing, Education and Language Studies
Centre for Research in Education and Educational Technology (CREET)
The Open University
September 2020

LEVERHULME
TRUST

iet
Institute of Educational
Technology



Abstract

This thesis reports on an analysis of an 11-million-word corpus of online discussions in a public educational site to understand the dialogic nature of online discourse. Two observations raise concerns that information exchange, socialization and online deliberation might be compromised in online spaces. Firstly, while anyone is free to express themselves online, they may not receive replies from others nor engage in sustained conversations with others. Secondly, afforded by the hyperlinking function, anyone can also easily share sources of information online by posting URLs, contributing to the circulation of (mis)information. This thesis thus explores how internet users can engage in meaningful dialogue with each other in online spaces through particular discourse and URL-posting practices.

A corpus linguistic approach comprising keyword analysis and micro-analysis is adopted to investigate how conversations are initiated and unfold within threads. Three keyword analyses are reported: (1) initiating posts, i.e., posts that receive replies; (2) independent posts, i.e., posts that do not receive any replies; and (3) the replies themselves. Based on these keyword analyses and informed by the theoretical concepts of dialogic space and intersubjectivity, linguistic features and discourse practices characterizing the two types of posts and replies are identified and explored. Similarly, URL-posting practices are also investigated to explore their role in online discourse.

Findings show that users draw on different discourse practices to invite replies and sustain conversations with others, although there are times users respond to the content on the site instead of addressing others. Importantly, discourse practices that do not entertain others' voices are found to deter others from responding or hinder their conversations, especially in the case of disagreement. In fact, disagreement provides an opportunity for users to explore different voices and achieve mutual understanding when discourse practices facilitative of intersubjectivity are utilized. Finally, although most users are positive towards URL-posting, the posting of URLs is seen to either facilitate or hamper their conversations, depending on users' posting and discourse practices.

Overall, this thesis highlights the role of various discourse practices in creating a dialogic space. A dialogic space allows multiple voices to be entertained in processes of intersubjectivity, such that users can engage with each other's subjectivities, whether they agree or disagree. Together, these findings highlight users' agency as enacted through language in online spaces and show that the discursive construction of online dialogic space should be one aspect of digital literacies of which internet users be made aware.

Acknowledgement

This thesis is not possible without Caroline who goes above and beyond the role of a supervisor. Thank you for always believing in me and encouraging me to have my own voice amid all the noises, and reading my numerous writing. I have enjoyed all the co-constructive sessions and some fun email exchange we have, and I hope you have too (since dialogic is the rule!). I have learnt a lot from you on writing, thinking, and be more creative and independent. Thanks also go to Bart and Mike, my other supervisors. Sarah from PACE has also been providing constructive and encouraging feedback for my draft to put me on track.

I am also grateful to Leverhulme Trust for funding this PhD journey and me attending numerous trainings and conferences. It is through these meetings that I learnt the inside-out of corpus linguistics and social sciences. Thanks also go to Paul Baker, Andrew Hardie and Stefan Everts for providing advices on the technicality of compiling and analysing a corpus.

This journey would not be a joyful one without my PhD comrades: particularly Pin who has volunteered to help me out so many times and listened to all my nonsense; Vasudha, Vicky and Quan who are the sensible ones; Dana who is my corpus linguistic buddy.

The task of writing becomes more enjoyable in Bogota with lovely staff and regulars. The stress of intensive writing during the pandemic is also lightened thanks to the help of Giri and the companionship of Pablo, while Caroline picked up the role of virtual support bubble. I also would like to thank Sue, Pek Ghe and other friends who have lent a hand whenever I need it. Lastly, I would like to thank my MK dad, Simon who saved me from a housing nightmare for my last lap of PhD. Finally, I would like to thank my two brothers and Bobby who listen to my rant every week, and my parents for putting up with the absence of their daughter.

Declaration of Authorship

I declare that this thesis is my own work. It has not been submitted for a degree or other qualification to the Open University or any other university or institution for examination.

Preliminary analysis arises from this thesis have been published in two conference proceedings:

Chua, Shi-Min. (2018). Why did Nobody Reply to Me? A Keyword Analysis of Initiating Posts and Lone Posts in Massive Open Online Courses (MOOCs) Discussions. In: *the 6th Conference on Computer-Mediated Communication (CMC) and Social Media Corpora*, 21-26.

Chua, Shi-Min; Tagg, Caroline; Sharples, Mike and Rienties, Bart. (2017). Discussion Analytics: Identifying Conversations and Social Learners in FutureLearn MOOCs. In: *FutureLearn data: what we currently have, what we are learning and how it is demonstrating learning in MOOCs*, 13-17 Mar 2017, Vancouver.

Table of Contents

Abstract	i
Acknowledgement.....	iii
Declaration of Authorship	iv
Table of Contents	v
List of Tables.....	xiii
List of Figures.....	xiv
List of Appendices.....	xvi
Glossary	xvii

Chapter 1

Introduction

1.1 Prelude	1
1.2 General aim of the thesis	3
1.3 Background to the thesis: Online discussions as sites of online discourse inquiry	4
1.3.1 Online discussions: Definition and observation	4
1.3.2 Online discourse and digital literacies: Motivations for the thesis	8
1.4 Research objectives	11
1.4.1 Empirical objective	11
1.4.2 Theoretical objective	12
1.4.3 Methodological objective.....	13
1.5 Research Questions	14
1.6 Outline of the thesis	15

Chapter 2

Research Setting:

MOOC online discussions

2.1 Introduction	19
2.2 MOOC.....	20
2.3 FutureLearn	21

2.4 Research on MOOC online discussions	28
2.4.1 Coding and counting.....	28
2.4.2 Corpus linguistic approach to MOOCs.....	30
2.5 Conclusion	32

Chapter 3

Research into online discussions and the dialogic nature of online discourse

3.1 Introduction	33
3.2 User-user interactions in online discussions	35
3.3 Evolution of threads and polylogical nature of threads	39
3.4 Linguistic features and content characteristics of posts	42
3.5 Discourse practices in replies	44
3.6 URL-posting in online spaces	46
3.7 Theoretical concepts for exploring dialogic nature of online discourse	53
3.7.1 Dialogic space	53
3.7.1.1 Dialogic language use and heteroglossia.....	54
3.7.1.2 Linguistic perspective	55
3.7.1.3 Educational perspective	57
3.7.1.4 Integrating two perspectives to apply dialogic space in online discussions.....	59
3.7.2 Intersubjectivity.....	59
3.7.2.1 Definition as integrated from education and linguistics	59
3.7.2.2 Stance-taking	61
3.7.2.3 Epistemic status and stance	62
3.7.2.4 Meta-language	63
3.7.2.5 Identity	65
3.8 Importance of sustained conversation for exploratory talk and intersubjectivity	67
3.9 Importance of disagreement for exploratory talk and intersubjectivity.....	68
3.10 Conclusion	70

Chapter 4

A corpus linguistic approach to online discourse

4.1 Introduction	73
4.2 Position towards textual contributions in online spaces	75
4.3 A corpus linguistic approach to online discourse.....	76

4.4 Corpus studies on online discourse	79
4.5 Concepts and methods in corpus linguistics.....	84
4.5.1 Corpus.....	84
4.5.2 Keyword Analysis.....	85
4.5.3 Statistical analysis for keyword analysis.....	88
4.5.4 Collocation analysis	91
4.5.5 Concordance.....	92
4.6 Applying corpus linguistics to online discussions in FutureLearn	93
4.7 Micro-analysis of discussion threads	95
4.7.1 Turn-taking in CA.....	95
4.7.2 Context in CA.....	97
4.7.3 Procedures in CA	98
4.8 Conclusion	100

Chapter 5

The FutureLearn Corpus

5.1 Introduction: The FL corpus.....	103
5.2 Corpus compilation.....	106
5.2.1 Data collection.....	106
5.2.2 Data processing	108
5.2.3 Corpus tool	112
5.2.4 Statistical tool and qualitative analysis tool.....	113
5.3 Ethical considerations	114
5.4 The FL Corpus: Further descriptions	117
5.4.1 MOOCs in the corpus.....	117
5.4.2 Users' comments and wordcount in each MOOC.....	121
5.4.3 Sub-corpora of users' comments for keyword analysis	122
5.4.4 Length of threads	124
5.4.5 Contributing users	128
5.4.5.1 Frequency of posting: Super-posters and one-time posters	129
5.4.5.2 Users' contribution patterns: Seven types of users	132
5.5 Conclusion	140

Chapter 6

A Keyword Analysis of Initiating Posts vs. Independent Posts:

Potential start of dialogic conversations

6.1 Introduction	143
6.2 Keyword Analysis: Initiating Posts vs. Independent Posts	144
6.2.1 Initiating keywords	144
6.2.2 Independent keywords.....	146
6.3 Interpretation of keywords: Functional grouping.....	148
6.3.1 Modals/modal Expressions	151
6.3.1.1 <i>I would</i> : Expressing interest or seeking information	152
6.3.1.2 <i>I would</i> : Positioning oneself.....	154
6.3.1.3 <i>it would</i> be + adjective: Evaluation as an introductory frame to stance or interest	156
6.3.1.4 <i>it would</i> and other instances of <i>would</i> : Prediction with uncertainty	157
6.3.2 Hedges	157
6.3.2.1 <i>perhaps</i> : Softening stance	158
6.3.2.2 <i>seems</i> : An introductory frame to stance	158
6.3.3 Concluding remarks on modals and hedges.....	159
6.3.4 Negative evaluative keywords: Admitting one's mistakes.....	160
6.3.5 Negation	161
6.3.5.1 <i>I don't know</i> : Information/help seeking	161
6.3.5.2 <i>I don't think/understand</i> : Stance expression.....	162
6.3.5.3 Concluding remarks on negation.....	166
6.3.6 Mental verbs: <i>wonder</i> and <i>wondering</i> to hedge questions and stance.....	167
6.3.7 Communicative verb forms	169
6.3.7.1 <i>say, tell</i> and <i>mean</i> : Stance expression.....	170
6.3.7.2 <i>says, told</i> and <i>called</i> : Intertextuality for evidentiality and information sharing	171
6.3.7.3 <i>tell, mean</i> and <i>explain</i> : Information/help seeking	172
6.3.8 Concluding remarks on mental verbs and communicative verbs	172
6.3.9 Meta-language on learning and discussion.....	173
6.3.9.1 <i>question</i> : Signposting information seeking	173
6.3.9.2 <i>article and question</i> : Intertextuality and information sharing	174
6.3.9.3 Concluding remarks on meta-language.....	174
6.3.10 Indefinite pronouns: <i>anyone</i> and <i>anybody</i> to call on others	175
6.3.10.1 Information/help seeking	175
6.3.10.2 Information sharing	176
6.3.10.3 Stance expression.....	176
6.3.10.4 Concluding remarks for indefinite pronouns.....	177
6.3.11 Polite speech-act formulae	177
6.3.11.1 <i>please</i> and <i>sorry</i> : Politely seeking information/help	178
6.3.11.2 <i>sorry</i> but I don't agree: Sticking to one's stance	178
6.3.11.3 <i>sorry</i> in meta-pragmatic expressions	178
6.3.11.4 Concluding remarks on polite speech-act formulae.....	179

6.3.12 Connectors	180
6.3.12.1 <i>if you</i> : Establishing interpersonal relationship	180
6.3.12.2 <i>if</i> -conditionals: Stance expression	181
6.3.12.3 Concluding remarks on connectors	182
6.3.13 Other functional groups of keywords	182
6.3.14 Uncategorized keywords: <i>example</i> and <i>e.g.</i>	183
6.3.14.1 <i>for example</i> : Linking to a typical case	184
6.3.14.2 <i>for example</i> : Parenthetical insertion	185
6.3.14.3 Adjective + <i>example</i> : Meta-language on learning and discussion.....	185
6.4 Discussion.....	190
6.4.1 General patterns of independent posts	191
6.4.2 General patterns of initiating posts.....	193
6.4.3 Discourse practices in initiating posts that are of dialogic nature	194
6.4.3.1 Indicating one's epistemic status: intention to know, "partially knowing", "unknowing"	194
6.4.3.2 Addressing potential readers to realise dialogic nature of online discourse	195
6.4.3.3 Intertextuality.....	196
6.4.3.4 Setting up a shared space with multiple voices.....	197
6.5 Conclusion	197

Chapter 7

Independent Posts:

Dialogic contraction and/or user-content interactions?

7.1 Introduction	199
7.2 Functional grouping of independent keywords: Expressions of appreciation.....	200
7.3 Analysis of selected keywords: <i>think</i> and <i>agree</i> for stance-taking.....	203
7.3.1 <i>I think</i>	204
7.3.1.1 Personal learning	205
7.3.1.2 Stance-taking	205
7.3.2 <i>I agree</i>	207
7.3.2.1 <i>I agree</i> with elaboration	207
7.3.2.2 Mere expression of agreement <i>I agree</i>	211
7.4 Discussion.....	213
7.5 Conclusion	216

Chapter 8

A keyword analysis of replies: Discourse practices for intersubjectivity

8.1 Introduction	219
8.2 Keyword analysis	220
8.3 Functional grouping of reply keywords.....	228
8.3.1 Stance-taking.....	229
8.3.2 Meta-language on others' comments.....	230
8.3.3 Interactive language.....	231
8.4 Conclusions regarding keyword analysis and the potential importance of disagreement	232
8.5 Micro-analysis of threads.....	234
8.5.1 "I agree"	234
8.5.2 "True but..." Concession strategy in negotiation.....	240
8.5.2.1 Context	240
8.5.2.2 Keywords	241
8.5.2.3 Start of the thread	241
8.5.2.4 Discourse practices in replies	242
8.5.2.5 End of the thread.....	245
8.5.3 "I did not say..." Meta-language and <i>if</i> -conditionals.....	246
8.5.3.1 Context	246
8.5.3.2 Keywords	247
8.5.3.3 Start of the thread	247
8.5.3.4 Discourse practices in replies	249
8.5.3.5 End of the thread.....	252
8.5.4 "You have never <i>replied</i> to my original <i>point</i> , but that would make a good academic discussion!" Metapragmatic discussion	254
8.5.4.1 Context	254
8.5.4.2 Keywords	254
8.5.4.3 Start of the thread	255
8.5.4.4 Discourse practices in replies	256
8.5.4.5 End of the thread.....	259
8.6 Conclusions regarding micro-analysis and the importance of <i>agree to disagree</i>	261
8.6.1 Concession with <i>but</i>	261
8.6.2 Meta-language for clarification and criticism	262
8.6.3 Metapragmatic expressions	263
8.6.4 Putting together keyword analysis and micro-analysis of threads	263
8.7 Agree to disagree.....	265
8.7.1 Interaction patterns in the threads	266
8.7.2 Acknowledging disagreement without negotiation	269

8.7.3 Reconciliation	272
8.7.4 Summarizing differences following negotiation.....	273
8.7.5 Framing on-going discussions.....	277
8.7.6 Failing to <i>agree to disagree</i>	279
8.7.7 Conclusions regarding <i>agree to disagree</i>	282
8.7.7.1 Disengage from discussion	283
8.7.7.2 Reconcile and maintain differences	283
8.7.7.3 Individuality and communicative norms in online discussions	284
8.8 Discussion.....	285
8.9 Conclusion	287

Chapter 9

Discourse practices of URL-posting

9.1 Introduction	290
9.2 Extent of URL-posting on FutureLearn	291
9.2.1 URL-posting varies across courses and types of comments.....	291
9.2.2 URL-posting varies across individuals.....	294
9.2.3 Major sources of URLs posted.....	296
9.3 General discourse patterns of URL-posting on FutureLearn	298
9.3.1 Collocation analysis	298
9.3.1.1 Collocates of URLs: how users include URLs in comments.....	298
9.3.1.2 Collocates of <i>link(s)</i> : how users respond to URLs	302
9.3.2 Length of comments containing URLs.....	304
9.3.2.1 Unaccompanied URLs	305
9.3.2.2 Minimal wording with the URLs	309
9.3.2.3 Long comments incorporating URLs.....	313
9.4 Conclusions regarding the extent and general discourse practices of URL-posting....	319
9.5 Micro-analysis	321
9.5.1 Co-constructing the value of Wikipedia as a source	322
9.5.1.1 “Thank GOD for <i>Wiki!</i> ”	322
9.5.1.2 “You should be wary of using <i>Wikipedia</i> ”	324
9.5.2 Using URLs to respond to URLs	327
9.5.2.1 “your one posted <i>link</i> ” vs “My <i>link</i> carefully explains”	328
9.5.2.2 Repeatedly posting the same URL.....	335
9.5.3 Superiority of URLs	340
9.5.3.1 “Can you back that up with some <i>links</i> ?”	340
9.5.3.2 “Unlike you, I’m willing to give a <i>link</i> ”	343
9.5.3.3 “We have provided <i>links</i> ”	348
9.5.4 Conclusions regarding micro-analysis	354
9.6 Conclusion	357

Chapter 10

Discussion and Conclusion

10.1 Introduction.....	359
10.2 Key findings	360
10.2.1 Prompt-focused posting.....	361
10.2.2 Discourse creates relationship among users in a dialogic space.....	363
10.2.3 Disagreement triggers sustained interactions as a dialogic space.....	370
10.2.4 Discourse practices as double-edged swords for engaging in intersubjectivity	371
10.2.5 URL-posting reflects users' evidencing practice.....	374
10.2.6 Summary.....	375
10.3 Empirical contributions	376
10.4 Theoretical contributions	377
10.4.1 Expanding the concepts of dialogic space and intersubjectivity.....	377
10.4.2 Empirical accounts of dialogic space and intersubjectivity.....	379
10.5 Methodological contributions	381
10.5.1 Dissecting online discussions.....	381
10.5.2 Integrating keyword analysis and micro-analysis.....	382
10.5.3 Corpus linguistics for analysing online discussions in MOOCs	383
10.6 Practical Implications	384
10.6.1 Discourse practices as one aspect of digital literacies	384
10.6.2 FutureLearn online discussions as a third space	387
10.7 Limitations and future research.....	388
10.7.1 Generalizability.....	389
10.7.2 Discourse as observed	390
10.7.3 Other discourse practices for dialogic conversations	392
10.8 Concluding Remarks.....	392
 References.....	 394
Appendix.....	412

List of Tables

Table 3.1 Selected studies that have reported quantitative information regarding user-user interactions	37
Table 3.2 Selected studies that have examined URL-posting in online discussions.	49
Table 4.1 Methods used in this thesis	74
Table 4.2 Selected corpus studies on online discourse	80
Table 4.3 The distribution of four words with different DP.	90
Table 5.1 The data access process with different universities	107
Table 5.2 Changes in the features of FutureLearn since the data collection for the corpus.....	108
Table 5.3 Data in the comment file that is to be processed and encoded into the corpus	109
Table 5.4 Summary of the 12 MOOCs included in the corpus.....	119
Table 5.5 Users' comments and wordcount across the 12 MOOCs.	121
Table 5.6 Number of comments and wordcount of the three types of comments in the corpus.	122
Table 5.7 Number of replies and wordcount of two types of replies in the corpus.....	124
Table 5.8 Super-posters contributions in each MOOCs.....	131
Table 6.1 Initiating keywords ordered by effect size.	145
Table 6.2 Independent keywords ordered by effect size.	147
Table 6.3 Functional grouping of the initiating keywords and independent keywords.	150
Table 6.4 Seven functional groups not elaborated in the text	183
Table 7.1 Functional grouping of independent keywords and initiating keywords	201
Table 8.1 Reply keywords that are used significantly more often when compared to both initiating posts and independent posts.....	222
Table 8.2 Reply keywords used significantly more often in first contributions and subsequent contributions in a thread.	225
Table 8.3 Reply keywords used significantly more often in short threads.....	226
Table 8.4 Reply keywords used significantly more often in the first reply of one-reply threads than those of threads with more than one reply.....	227
Table 8.5 Functional grouping of the reply keywords.	229
Table 8.6 Normalized frequency in per million words of agree to disagree and agree to differ in different corpus	266
Table 8.7 Interaction patterns in the threads where agree to disagree/differ is used.	267
Table 9.1 URLs posted in each type of comment in each course	292
Table 9.2 Major sources of URL posted.	296

List of Figures

Figure 2.1 One course step that contains a cartoon and an article.....	23
Figure 2.2 The commenting space below the course content in each step.	24
Figure 2.3 One course step that contains mainly discussion prompts.	26
Figure 2.4 One course step that contains a video and discussion prompts.	27
Figure 3.1 How hyperlinking to URLs is shown in FutureLearn online discussions.	47
Figure 4.1 Concordance of the word “FutureLearn” in the corpus.	92
Figure 4.2 Summary of methodology in this thesis.	101
Figure 5.1 Components of the corpus compiled and examined in this thesis.....	105
Figure 5.2 Data-processing	111
Figure 5.3 Nested structure of online discussions in FutureLearn	113
Figure 5.4 Number of threads of different length across the 12 MOOCs.....	125
Figure 5.5 Scatterplot: Number of subsequent contributions vs. length of threads.....	126
Figure 5.6 Scatterplot: Number of users involved vs. length of threads.....	128
Figure 5.7 Comparison of users’ contribution across the 12 MOOCs.....	130
Figure 5.8 Each user’s contribution of new posts and replies	133
Figure 5.9 Different types of users categorized based on their contributions.....	135
Figure 5.10 Proportion of comments contributed by different types of users compared to the proportion of each type of users.	139
Figure 6.1 Thread 5157379.	153
Figure 6.2 Thread 4954168.	155
Figure 6.3 Thread 18994217	163
Figure 6.4 Thread 19605888	165
Figure 6.5 Thread 4329045	168
Figure 6.6 Concordance lines of “for example” used as parenthetical expression.	185
Figure 6.7 Thread 4442056	187
Figure 6.8 Thread 19531657	188
Figure 6.9 Thread 19347934	189
Figure 7.1 Independent post 4626201.....	206
Figure 7.2 Independent post 17474926.....	206
Figure 7.3 .Expression of agreement towards course content with elaboration.	209
Figure 7.4 Expression of agreement towards a specific user with elaboration.....	209
Figure 7.5 Expression of agreement with elaboration towards comments contributed by other users.	209
Figure 7.6 “I agree” may be used to respond to the “Do you agree” discussion prompt.	210
Figure 7.7 Mere expression of agreement towards course content.	211
Figure 7.8 Mere expression of agreement towards course content.	212
Figure 7.9 Mere expression of agreement towards comments contributed by others.	212
Figure 7.10 Mere expression of agreement towards comments contributed by others.	212
Figure 8.1 Expression of agreement in replies.....	235
Figure 8.2 Expression of agreement in replies with additional content.....	236
Figure 8.3 Expression of agreement in replies in a long thread.	238
Figure 8.4 Start of thread 18016862.....	241
Figure 8.5 Reply 2 to reply 6 of thread 18016862 where concession practices are observed.....	243
Figure 8.6 End of thread 18016862	245

Figure 8.7 Start of thread 19605035.....	248
Figure 8.8 Reply 2 to reply 5 of thread 19605035 where qualification and meta-language are observed.	250
Figure 8.9 End of thread 19605035	253
Figure 8.10 Start of thread 20392679.....	255
Figure 8.11 Reply 2 to reply 5 of thread 20392679 where metapragmatic discussion happens	257
Figure 8.12 End of thread 20392679	260
Figure 8.13 Part of thread 4352189.....	270
Figure 8.14 Part of thread 4505886.....	272
Figure 8.15 Part of thread 18030898.....	274
Figure 8.16 Part of thread 19422436.....	276
Figure 8.17 Part of thread 19180872.....	278
Figure 8.18 Part of thread 19412152.....	280
Figure 9.1 Scatterplot: Number of URLs posted vs. Number of comments posted.	295
Figure 9.2 Thread 18971576.....	303
Figure 9.3 Number of words in the comments containing URLs.....	304
Figure 9.4 Unaccompanied URL in an independent post	305
Figure 9.5 Unaccompanied URL in an initiating post.....	306
Figure 9.6 An unaccompanied URL in reply 8	308
Figure 9.7 Minimal introduction of URLs in the initiating post and reply 2.....	311
Figure 9.8 The URL is introduced in length in the initiating post.....	314
Figure 9.9 Two URLs are recommended in reply 1 where user writes about how useful they are. ..	315
Figure 9.10 The URL is cited as “an example”	317
Figure 9.11 Thread 4423693. The text of the URL is quoted verbatim.	318
Figure 9.12 All the replying users thank the URL-posting user for sharing the Wikipedia link.	323
Figure 9.13 Discussion about Wikipedia	325
Figure 9.14 First part of thread 20311486.....	329
Figure 9.15 Second part of thread 20311486	330
Figure 9.16 The characteristic of the URL-URL interaction in thread 20311486.	332
Figure 9.17 First part of thread 18980719.....	336
Figure 9.18 Second part of thread 18980719	337
Figure 9.19 Thread 6400817.....	341
Figure 9.20 First part of thread 703441	344
Figure 9.21 Second part of thread 703441	345
Figure 9.22 User ah1-993’s replies that introduce or quote URLs.....	349
Figure 9.23 user ah1-12’s replies that are about their personal experience.	350
Figure 9.24 Some of user ah1-993’s replies that reveal the importance of URLs to them.	351
Figure 9.25 Some of user ah1-12’s replies that reveal their negative sentiments towards URLs.	352
Figure 10.1 Word cloud of initiating keywords.....	365
Figure 10.2 Word cloud of independent keywords.	367
Figure 10.3 Word cloud of reply keywords.....	373

List of Appendices

Appendix A Facilitators' comments and wordcount in each MOOC	412
Appendix B Ethical approval from the Open University Human Research Ethics Committee	413
Appendix C Statistics for keyword analysis of initiating posts compared to independent posts	414
Appendix D Statistics for keyword analysis of independent posts compared to initiating posts	416
Appendix E Number of likes received by independent posts	418
Appendix F Statistics for keyword analysis of replies when compared to initiating posts.....	419
Appendix G Statistics for keyword analysis of replies when compared to independent posts.....	421
Appendix H Statistics for keyword analysis of reply keywords comparing first contributions and subsequent contributions.....	425
Appendix I Statistics for keyword analysis of reply keywords comparing short threads and long threads.	426
Appendix J Statistics for keyword analysis of reply keywords comparing first reply of one-reply threads and that of threads with more than one reply.....	427

Glossary

Online discussion	An online space where users can post their comments for others to see and reply to.
Comment	Generic term used to refer to users' posting in online discussions. It refers to both post and reply.
New post/post	A comment that is posted separately from others' comments. Each new post has the potential to elicit replies from other users and initiates a thread of discussion.
Initiating post	A post that receives replies from other users and thus initiates a thread of discussion.
Initiator	The user who contributes the initiating post of a thread, thus initiating the thread.
Independent post	A post that does not receive replies from other users.
Reply	A comment underneath an initiating post.
Initiating keyword	A word that is used significantly more often statistically in initiating posts compared to in independent posts.
Independent keyword	A word that is used significantly more often statistically in independent posts compared to in initiating posts.
Reply keyword	A word that is used significantly more often statistically in replies compared to in initiating posts and independent posts.
Thread	A thread is formed by an initiating post and replies.
Short thread	A thread with four or fewer than four replies.
Long thread	A thread with five or more than five replies.
First contribution	The first time a user replies in a thread.
Subsequent contribution	The reply contributed by a user who has already replied in the thread before, or who has initiated the thread.
MOOC	Massive Open Online Course.
Futurelearn	A MOOC platform that is the research setting of this thesis.

Course step	In FutureLearn, a course step is equivalent to a page on a website. Most steps contain a discussion space.
Discussion prompt	In some steps on FutureLearn, discussion prompt is used to invite users to respond in the discussion space.
Prompt	In online discussions in general, prompt is referred to the content on the web page, including video, audio, picture, text etc.
User	A person who joins the online space.
Learner	A user who joins the MOOC to learn.
Facilitator	An expert whose role in the MOOC online discussion is to intervene and teach.
Super-poster	A user/learner who comments a lot in an online space.
One-time contributor	A user who only comments once.

Chapter 1

Introduction

This introductory chapter gives an overview of the research aim and scope of the thesis.

1.1 Prelude

Post 1 (posted on 9 April 2015, liked by 8 users)

“was bemused by the cartoon at the beinning...ok so you may not like what i write but noticed the 'rich' family had one child and the 'poor' had two...is this part of the inequality in society that some and i repeat some people have children they cannot afford but expect someone to pick up the tab by having more benefits such as tax credits child benefit needing larger houses etc ..it is just a question..family and friends who have more children are generally poorer”

Post 2 (posted on 8 April 2015, liked by 3 users)

“Should we be looking at the whole system rather than blaming the baby boomers for everything?
<http://www.theguardian.com/commentisfree/2015/apr/08/rising-inequality-technological-change-loss-jobs>
Professor Anthony Atkinson has a lot of good points - and the current system might - probably was - designed to produce, preserve and increase inequality. And there is an election on 7th May in the UK.”

Post 3 (posted on 17 April 2015, liked by 8 users)

“Have only managed to start week 4 today, Saturday, but am really saddened by some of this discussion. I've always been proud to be a taxpayer, proud that these are my roads, my hospitals, my teaching staff etc etc ... and also proud that I / we can support those who need it.”

The above three posts come from the same online discussion on a content webpage titled “Pension, housing and wider inequalities”. Typos and misspellings are retained from the original posts. Where these posts come from will be revealed and described in Chapter 2. Other users can reply to these posts and start a conversation, although these posts can also remain not replied to. Which post would other users more likely reply to?

Despite the fact that all these posts were seen by other users, as evinced by the number of “likes” they received, and can be considered as on-topic in terms of their content, Post 1 received 51 replies, thus triggering a long thread, and Post 2 only received one reply, thus a short-lived discussion. In contrast, Post 3 received none, thus not starting a thread at all. Why is there such a difference between these posts? It may be chance. For example, Post 3 was posted more than one week after the other two posts, such that there might be fewer people seeing it. However, in this example, Post 1 and Post 3 both received eight “likes”, suggesting that both have been read by at least eight other users. Also, Post 1 and Post 2 were posted just one day apart and were ordered next to each other in the discussion space, yet there is a huge difference in the length of threads generated. In fact, the effect of time of posting in receiving replies is inconclusive at best. Some studies found no evidence (Arguello et al., 2006; Rooderkerk & Pauwels, 2016), while Hewitt (2003) found the more recent posts and Jeong & Frazier (2008) found the earliest posts are more likely to receive replies.

Therefore, the observation that one post can receive so many replies when another goes unanswered raises the question as to whether there is anything in the post itself, or the *language* used, that either attracts other users to respond or deters them from doing so upon reading the post. The huge difference in the length of discussion threads generated by Post 1 and Post 2 also raises the question whether the language in the replies keeps a thread alive or kills it. In other words, the proposition that how a post is written might determine whether it is responded to suggests that internet users can exploit language to differential effects in online discussions.

Finally, besides language, it is evident that other semiotic resources may also play a role. One URL linked to an article in The Guardian is observed in Post 2. The user cited “Professor Anthony Atkinson” mentioned in the linked article and highlighted that the expert “has a lot of good points”. This suggests that the user shared and used this URL as a resource to support their stance. The URL-posting is made possible by the technological affordance of hyperlinking in most online spaces (Kiernan, 2018; Tyrkko, 2010). Importantly, URL-posting mirrors the consumption and (re-)sharing of

information in social media and other online spaces, that potentially involves the circulation of misinformation, polarized opinions, or scams (Secker, 2017; Tagg & Seargeant, 2019).

1.2 General aim of the thesis

The general aim of this thesis is to investigate the dialogic nature of online discourse in order to raise internet users' awareness of their discourse practices in establishing social interactions and engaging in deliberation in online spaces. *Discourse practice* refers to the way language is used within a particular social situation, in this thesis online discussions. *Dialogic* refers to the interplay of different voices in communication, in contrast to *monologic* communication (Bakhtin, 1981). A conversation can either be *monologic* or *dialogic* depending on whether each conversational partner entertains each other's voices. It has been observed that online discussions are sometimes not dialogic (Beth, Jordan, Schallert, Reed, & Kim, 2015; Meyer et al., 2019; Napoles, Tetreault, Rosato, Provenzale, & Pappu, 2017), as evidenced by users' posts left unresponded to, or short-lived conversations, as illustrated above. Even when users reply to each other, the conversation may be monologic rather than dialogic if they do not engage with each other's voice. Therefore, this thesis focuses on the discourse practices that users can employ to initiate a conversation with others; that is, attract replies from others, and engage in a dialogic conversation. The theoretical definition of *dialogic conversation* will be explained in section 1.4.2.

The investigation of this thesis is based on the assumption that language as a meaning-making tool is used by humans to enact their identity, construe their social world, and co-construct relationships with others (Fairclough, 2003; Heritage, 2012; Herring, 2004; Vygotsky, 1978). As such, the analysis of discourse practices in this thesis is not about *what* a post or a reply is, for example a question or a disagreement. Instead, this thesis is about *how*, for example, a question is asked, or a disagreement is raised to establish conversations with others. To achieve this aim, a large corpus of 11-million words of online discussions of Massive Open Online Courses (MOOCs, to be introduced in

Chapter 2) was compiled and analyzed. Corpus linguistics is a mixed methodology that exploits the advantages of big data by identifying repeated language patterns in a corpus and also allows detailed discourse analysis at the same time. The corpus analysis in this thesis sheds light on the way in which discourse is used to initiate and sustain online interactions in the form of broad patterns across the data and in the immediate context of an unfolding exchange among users within a thread.

1.3 Background to the thesis: Online discussions as sites of online discourse inquiry

The rise of internet means that users who do not know each other in person or are located at different corners of the world can connect with each other around a shared interest (e.g. pension inequality) or activity (e.g. learning) at any time for information exchange, emotional support and deliberation of issues. One such space is online discussion which, given its interactive nature, is suitable for exploring the dialogic nature of online discourse.

1.3.1 Online discussions: Definition and observation

In principle, online discussions can happen in any online space as long as the design of a website allows for users' contributions and interactions among multiple users. Online discussions can take place synchronously or asynchronously via various media including text, video, or voice.

Nonetheless, text-based asynchronous interactions remain the main means of social interactions and information flow among users, probably because users do not need to be engaged simultaneously (Herring, 2004; Jones & Hafner, 2012). This thesis focuses on text-based online discussions for the investigation of the dialogic nature of online discourse.

Online discussions are not only found in traditional online spaces dedicated for discussions, such as forums for Q&A and support/interest groups, newsgroups, bulletin board systems, mailing lists and social networking sites, but also in the commenting space alongside multimedia content such as in blogs, news-websites, TED, Facebook or YouTube (Arguello et al., 2006; Bolander, 2013;

Bou-Franch & Garcés-Conejos Blitvich, 2014; Frobenius & Harper, 2015; Herring, 2004; Jaworska, 2018; Meyer et al., 2019; Savolainen, 2014). Messaging apps, such as WhatsApp and Facebook Messenger, are also used for discussions, especially when a group is created, although in these cases, the users most likely know each other in person (Manipuspika, 2019).

Online discussions have multiple functions for both users and those hosting them. Users participate to seek or provide advice, support or information (Burke, Joyce, Kim, Anand, & Kraut, 2007; Connor, 2013; Jaworska, 2018), engage in deliberation with others (Dahlberg, 2001) and professional networking (Fayard & DeSanctis, 2005), or merely socialize for recreational purposes (Herring, 2001; Khan, 2017; Pendry & Salvatore, 2015; Springer, Engelmann, & Pfaffinger, 2015). Organizations or individuals host online discussions for public or customer engagement (Grabill & Pigg, 2012; Polletta, Chen, & Anderson, 2009; Wright & Street, 2007), and educators use online discussions to engage their students (Laurillard, 2012). Based on the forms and functions observed in the literature, the following working definition of online discussions is used in this thesis:

Online discussions are where user-content and user-user interactions take place, the latter of which involves information exchange, socialization and deliberation.

User-content interactions refer to users posting in response to content, especially in the commenting spaces of webpages that contain multimedia content. User-user interactions refer to users replying to others, or posting to invite others' response, and can take place in discussion forums or commenting spaces (Ksiazek & Lessard, 2016). User-user interactions can be where a conversation arises given that it is where users' voices are heard and replied to by others. Both user-content and user-user interactions are not only limited to textual contributions, but also non-verbal actions such as clicking and scrolling. User-user interactions also involve liking other users' contributions and following other users. However, this kind of user-user interactions does not open up a conversation. Therefore, this thesis only examines user-content and user-user interactions in

online discussions in terms of textual contributions to explore the dialogic nature of online discourse.

From the point of view of user-content interactions and users' engagement, a massive number of users' contributions in online discussions is considered a success (Preece & Maloney-Krichmar, 2005; Sharples & Ferguson, 2019). However, it is commonly observed that most posts in online discussions do not receive replies or only trigger short-lived threads, which can potentially compromise user-user interactions and the building of virtual communities (Bou-Franch & Garcés-Conejos Blitvich, 2014; Herring, 1999, 2013; Ksiazek & Lessard, 2016). Users in online discussions have expressed dissatisfaction and a reduced sense of community, or even given up posting, for lack of responses from others (Delahunty, 2018; Hewings, Coffin, & North, 2009; Joyce & Kraut, 2006; Springer et al., 2015), whereas Arguello et al. (2006) find that receiving a reply increase the chance of posting again for newcomers. Not receiving replies can be especially problematic when users are actively seeking emotional or informational support (Crook et al., 2016).

Furthermore, online deliberation requires users to listen to and reply to each other's arguments; that is, *dialogic conversation*, rather than just posting without engaging with others (Freelon, 2015; Friess & Eilders, 2015). The latter can potentially lead to polarization of views or echo chambers (Walter, Brüggemann, & Engesser, 2018). To differentiate these two scenarios, *online deliberation* is used in this thesis to specifically refer to online discussions where users deliberate on issues with different views and engage in dialogic conversations (Lewiński, 2013), while *online discussions* are general situations where users can simply post and do not necessarily engage with others' views. It should be acknowledged that online deliberation itself is a research topic in the politics literature that deals with using online spaces for public participation in decision making for reaching consensus in a democracy (Friess & Eilders, 2015). Nonetheless, the central idea of online deliberation – “dialogue and difference” (Dahlberg, 2001, p.616) is maintained in this thesis.

The fact that user-user interactions are often fewer than desired can be associated with the technological affordances of asynchronous text-based online discussions. Firstly, online discussions

create a levelling ground for users such that they could create their own posts or just write in response to the content of the web page rather than being obliged to engage with others (Bou-Franch, 2012; Cavanagh, 2007). Herring (2013) calls this *prompt-focused* posting when users comment mainly in response to the prompt on the page, i.e., the multimedia content on the page, rather than to other users' comments. Responding to prompts is an indication of user-content interaction. However, if most users engage in this prompt-focused posting, it might undermine user-user interactions which require users to respond to others instead of prompts. In turn, the large number of comments may further overwhelm users, and render each post and thread less likely to be responded to (Himmelboim, 2008).

Secondly, unlike face-to-face or other synchronous online conversations, users are not obligated to respond or continue a conversation, and do not even need to express leave-taking from an on-going conversation (Jones & Hafner, 2012). Thirdly, the asynchronous nature also makes it hardly possible to predict the best time to post to increase the chance of user-user interactions, as shown by the inconclusive evidence on the effect of time of posting (Arguello et al., 2006; Rooderkerk & Pauwels, 2016). Additionally, although users may strive to post at a time predicted to be more likely seen by others, this might be beyond individual users' control, given that online discussions are joined by users from different time-zones.

To rectify these internet-mediated factors related to lack of response to users' comments and to short-lived interactions, some researchers have proposed automatic recommendation of quality comments to be presented to users (Coetzee, Fox, Hearst, & Hartmann, 2014; Wise, Cui, & Vytasek, 2016). However, this recommendation approach might risk users focusing only on selected posts or threads to the extent of creating echo chambers, and marginalize users whose comments are not selected (Dron, 2014). Furthermore, exploration of an additional topic that is not recognized by the system might be compromised. To some extent, this approach borders on technological determinism which assumes that technology can determine social change (Herring, 2004).

1.3.2 Online discourse and digital literacies: Motivations for the thesis

Admittedly, technological design can potentially improve users' experience. However, in the era of technology, it is important not to overlook human agency in the digital world (Herring, 2004; Jones & Hafner, 2012; Tagg & Seargeant, 2019). One such agency can be enacted via language (Duranti, 2004). This concept of agency not only draws attention to users' discourse practices online, but also acknowledges what users can do within the technological affordances and constraints of online communications.

For example, asynchronous text-based online discussions afford multiple users contributing to the same thread at different times. However, without a threading system that indicates the relationship between comments within the thread, users could potentially be confused with which message is addressed to which message. To overcome this confusion, users use vocatives, quotes or meta-language to refer to the message they are addressing (Bolander, 2012; Bou-Franch, 2012; Herring, 1999; Thomas, 2002). This shows that, although under technological constraints, users can execute their agency in online spaces with their discourse practices.

More importantly, internet users can employ various discourse practices to enact their identity and social relationships with others, while their communication is mediated by the technological affordances and constraints of online spaces (Jones & Hafner, 2012). For example, to avoid or repair miscommunication that can jeopardize social relationships, users can comment on their own comments to offer further information on their previous comments (e.g., "I apologize if my post sounded somewhat condescending, because that was not the intent", Tanskanen, 2007, p.93). Newcomers to a group forum can become unwelcome when their discourse practices enact themselves as out-group rather than in-group and do not acknowledge the advice provided by the existing members (Stommel & Koole, 2010). In every online discussion and virtual community, the norms of discourse practices and social relationships are co-constructed by users (Beers-Fägersten, 2008; Herring, 2004; Tagg & Seargeant, 2019; Tanskanen, 2007). Consequently, users need to be

aware of the effect of their own and others' discourse practices in a particular online discussion or community.

Therefore, not only technology design but also users' discourse practices are important in creating an online space for information exchange, socialization and deliberation. Numerous researchers have argued for users to be made aware of how to engage with each other in online discussions, especially for deliberation when they hold opposing stances or polarizing views (Bou-Franch & Garcés-Conejos Blitvich, 2014; Freelon, 2015; Laflen & Fiorenza, 2012; Littleton & Whitelock, 2005; Paulus, 2006). For example, in one online deliberation space, second person pronouns are employed in discourse practices to admonish others or exclude others (Sotillo & Wang-Gempp, 2016). In such cases, although the use of second person pronouns generally indicates interactivity in online space (D. Knight, Adolphs, & Carter, 2014; Yates, 1996), it could also undermine social relationships in online deliberation. Therefore, awareness of such varied language usage is important for effective communication via digital media.

Users' discourse practices to enact their identity and social relationship with others is one aspect of digital literacies (Jones & Hafner, 2012; Thorne, 2013). Broadly speaking in the realm of digital literacies, this aspect can be considered as being skilled at presenting information or content to others with the right expressions, thus a social practice in online spaces (Lankshear & Knobel, 2006). This aspect of digital literacies lies on the production, as opposed to the consumption of information, which is one of the earlier conceptions of digital literacies by Gilster (1997). In recent years, educational institutions and policy makers have been promoting digital literacies, given that digital devices and internet are becoming ubiquitous and rapidly evolving, for good or ill. Generally, digital literacies entail skills and competencies of using digital tools and media, but there is hardly a consensus on what constitute digital literacies (Secker, 2017; Thorne, 2013). More nuanced digital literacies include awareness of algorithms behind search engines and social media, critical evaluation of information online, managing one's internet footprints, non-linear reading of hyperlinked texts,

engaging in online communities and collaborations. This list is not exhaustive, and changes as digital technology develops.

Although it is not possible to cover every aspect of digital literacies, this thesis explores how internet users exploit language in online spaces to establish social relationships, specifically, how they initiate and engage in a discussion thread. This aspect of digital literacies is crucial for users who expect responses from others or like to engage in conversations with others in online discussions. Furthermore, although online spaces afford users to freely express their views, to prevent polarization of views and echo chambers, it is of utmost importance for users to engage in online deliberation with others to build on each other's voices for a dialogic conversation (Dahlberg, 2001; Freelon, 2015; Friess & Eilders, 2015).

As mentioned in the prelude, URLs are another semiotic resource users employ for sharing information and supporting their arguments in online spaces (Connor, 2013; Sudau et al., 2014; Wikgren, 2001). URL-posting is an indication of users executing their agency in the digital world as an information distributor (Edgerly, Thorson, Bighash, & Hannah, 2016; Oeldorf-Hirsch & Sundar, 2015). However, there has been growing concern regarding the circulation of misinformation and polarization via URL-posting by internet users in online spaces such as social networking sites, blogs and online discussions (Burnett & Jaeger, 2008; de Maeyer, 2013; Tagg & Seargeant, 2019). What underlies users' posting of URLs may be related to how they evaluate and use information online, which is one aspect of digital literacies (Gilster, 1997; Lankshear & Knobel, 2006). Although researchers have begun to investigate URL-posting, the research is mainly on what types of URLs (e.g., academic sources, news media, commercial websites) are shared, with limited investigation on *how* URLs are used. This thesis thus fills this gap by examining discourse practices of URL-posting in online discussions, thereby raising users' awareness of their digital literacies in using and consuming URLs. The exploration of URL-posting also expands the investigation of discourse practices in this thesis to a semiotic resource other than language.

1.4 Research objectives

This thesis has three main objectives: empirical, theoretical and methodological. The empirical objective builds on what has been introduced thus far in this chapter on online discussions. The theoretical objective regards the dialogic nature of online discourse. The methodological objective is in relation to the online discussions to be examined, MOOCs. The objectives are as follows:

1. Empirical objective:

To explore the discourse practices employed by users in online discussions, including those that can initiate and sustain a conversation, and URL-posting.

2. Theoretical objective:

To provide the textual evidence for the theoretical concepts of dialogic space and intersubjectivity to characterize the dialogic nature of online discourse.

3. Methodological objective:

To exploit the big data available from MOOC online discussions with corpus linguistic approach and the view that language construes social relationship.

1.4.1 Empirical objective

As mentioned above, some users join online discussions to seek emotional or informational support, socialize or engage in deliberation with others. However, they do not necessarily receive replies from others, and there are times that the conversations are too short for meaningful discussion. Although several studies have begun to examine linguistic features that can increase the chance of receiving replies, these studies do not go into detail on how these features realize discourse practices for establishing conversations (e.g., Arguello et al., 2006; Crook et al., 2016; see Chapter 3 for details). This thesis thus seeks to unravel discourse practices that can attract replies and sustain interactions with others, so as to improve the experience of user-user interactions and promote dialogic conversations online.

In terms of real-world implications, this thesis aims to raise users' awareness of their agency and discourse practices in enacting social relationships with others in online spaces which bring both technological affordances and constraints to our communication. This can be considered as one of the digital literacies much needed in the era of internet where everyone can post freely. One specific discourse practice, URL-posting is also examined to further explore users' consumption and sharing of (mis)information online, another area of digital literacies.

1.4.2 Theoretical objective

The main theoretical objective is to provide textual evidence – linguistic features and discourse practices – that characterize the dialogic nature of online discourse, which might be most probably found in the online discussion threads. This is because initiating posts and replies are where users' dialogue, or polylogue happens, whereas independent posts may not be dialogic in nature to invite replies.

However, even within a thread or a face-to-face conversation, a true *dialogue* may not necessarily happen. Rather, parallel monologues can take place if users do not take into account or invite each other's voices in their conversations. This might mirror the polarized discourse and monologues seen in online spaces where people simply express their own views without attempting to listen and relate to others (Freelon, 2015). Therefore, although user-user interactions are where conversations occur, it is not sufficient for dialogic conversations. In this thesis, a *dialogic conversation* is defined as a conversation where all conversational partners' voices are entertained and made relevant to each other's (Du Bois, 2007; Martin & White, 2005). A dialogic conversation requires dialogic space to be established by users' discourse (Martin & White, 2005), while users also engage in the processes of intersubjectivity, i.e., the sharing and integration of subjectivity of each participating user (Du Bois, 2007; Hall, 2010; Stahl, 2015). It is worth bearing in mind that in this thesis, user-user interactions are used to refer to all the threads where users reply to the initiating posts or other replies, whereas dialogic conversations are used to characterize threads in which

users' discourse realize dialogic space and intersubjectivity. It is hoped that analysis of the linguistic features and discourse practices in each types of comments may enrich our understanding of the dialogic nature of online discourse, thus informing users of useful discourse practices for engaging in meaningful conversations online.

1.4.3 Methodological objective

The methodological objective mainly applies to the research setting of this thesis, the online discussions in MOOCs, a public educational space, which is to be introduced in Chapter 2. In short, research in MOOCs have mainly treated the textual contributions by users in the online discussions as representations of users' thinking, so users' comments are evaluated for the quality of discussions or individuals' learning outcome (Kellogg, Booth, & Oliver, 2014; Wang, Wen, & Rosé, 2016). To enhance user-user interactions in MOOC discussions, researchers mainly focus on developing algorithms to automatically organize and recommend comments to users, probably because of the educational and technological nature of MOOC online discussions (Almatrafi & Johri, 2019; Wise et al., 2016). Instead, this thesis approaches the online discussions from the point of view of users' agency in using language to do things, thus bringing in another methodological perspective to MOOC research. More importantly, this thesis aims to increase users' awareness of their discourse practices in this online educational space.

Another methodological objective is to apply corpus linguistic analysis, a mixed-methodology that draws on big data and qualitative discourse analysis, in MOOC research. The lack of detailed analysis of discourse practices in MOOC online discussions is probably also due in part to the massive number of discussion postings. MOOC researchers mainly resort to a coding and counting paradigm in which users' textual contributions are reduced to a limited number of codes, either manually or automatically, for counting purposes (see Chapter 2 for details). This thesis seeks to show that it is possible to examine discourse practices in big language data, in this case an 11-

million-word corpus, with a corpus linguistic approach without sacrificing the richness of the textual data.

1.5 Research Questions

In this thesis, I will focus on how users enact their agency with their language use to establish dialogic conversations with others in online discussions for information exchange, socialization and deliberation. Specifically, I will examine how users can initiate and sustain conversations with others in the face of lack of responses from others and short-lived threads. The research questions to be addressed are:

RQ1: What are the differences in the linguistic features and discourse practices that regularly occur in

- **initiating posts that receive replies and start a discussion thread,**
- **independent posts that do not receive replies,**
- **replies, especially those in sustained discussions?**

RQ2: How do these discourse practices initiate, sustain or hinder dialogic conversations in online discussions?

RQ3: How does URL-posting initiate, sustain or hinder dialogic conversations in online discussions?

The comments in online discussions are first dissected into three types: (1) initiating posts that receive replies and start a thread, such as Post 1 and 2 illustrated above; (2) independent posts that receive no replies and thus not in a thread, such as Post 3; and (3) replies within a thread. The differentiation allows a more fine-grained analysis of the dialogic nature of user-user interactions, rather than treating all comments as the same. It also paves the way for the quantitative analysis of

the corpus linguistic approach that reveals *keywords*, i.e., words used statistically significantly more often in each type of comment. A qualitative discourse analysis of these keywords and micro-analysis of threads reveal different discourse practices that users can employ to initiate and engage in conversations, thus revealing the dialogic nature of online discourse.

To address the research questions, *linguistic features* and *discourse practices* in the three types of comment will be derived from the keywords found. Following Biber, Johansson, Leech, Conrad, & Finegan (1999) and Myers (2010), *linguistic features* are used loosely in this thesis to refer to any object of inquiry for linguistic analysis. They could be grammatical structures, lexical choices, i.e., words, or lexical collocations, i.e., phrases, that users use to achieve their communication, for example, the use of “X” in text-messaging (Tagg, 2012), or modality, evaluation, hedging, and meta-language (Biber et al., 1999; Carter & McCarthy, 2006). Linguistic features typically are used to refer to language units smaller than a sentence. In contrast, *discourse practices* are above and beyond a sentence, and in this thesis refer loosely to the recurring ways a language community use language to do things in social context (Fairclough, 2003). Discourse practices can change from one community to another, and can be actualized by different linguistic features in different social situations. For example, *small confession*, as Jaworska (2018) terms it, is a discourse practice employed by mothers suffering post-natal depression in Mumsnet to carve out a digital space for shared experience and for self-empowering rather than self-pitying. Simply put, discourse practices are social behaviours that are observable based on textual evidence, i.e., language and linguistic features.

1.6 Outline of the thesis

Chapter 2 introduces MOOCs, the research setting, or rather the online space for the examination of online discourse in this thesis. It also reviews the educational research on users’ textual

contributions in the online discussions in MOOCs, and points to the lack of research on understanding users' discourse practices in this online educational space.

Chapter 3 reviews literature on user-user interactions, online discourse and URL-posting. This review highlights the lack of in-depth analysis of discourse practices needed to understand how dialogic conversations can be initiated and sustained. Two theoretical concepts, dialogic space and intersubjectivity, are explained for their potential in directing the inquiry of the dialogic nature of online discourse. These two concepts also distinguish dialogic conversations from user-user interactions.

Chapter 4 introduces and justifies the corpus linguistic approach adopted in this study. It explains the novel approach of distinguishing between initiating posts, independent posts and replies as the basis for keyword analysis. Linguistic features and discourse practices of each type of comment are then derived from the keywords with corpus methods. Micro-analyses are then carried out to examine the text of the initiating post and all the replies within selected threads to understand how each comment builds on the previous comments and affect the next one. By combining corpus methods with micro-analysis, the analysis capture both the general patterns and the nuances of users' discourse in the online discussions.

Chapter 5 outlines the compilation of the 11-million-word corpus which consists of 12 MOOCs. The corpus is described in terms of users' posting behaviours to provide background to the users' discourse in this online space. The investigation into online discourse are presented in Chapter 6 to 9. Each chapter starts with quantitative analysis to give an overview and general patterns of users' discourse before going into detailed qualitative analysis of specific linguistic features or discourse practices.

Chapter 6 begins with a keyword analysis comparing initiating posts and independent posts, both of which have the potential to start dialogic conversations. The keyword analysis reveals differences in the linguistic features and discourse practices between these two types of posts. It is

argued that discourse practices in the initiating posts, as opposed to independent posts, are more likely to invite replies from others by creating a dialogic space and establishing intersubjective relationships. This is the first set of empirical findings in this thesis that shows that users' agency executed via their discourse practices in online discussions matters for the success of their interactions, in this case in inviting replies.

In Chapter 7, it is argued, although discourse practices in independent posts may not establish dialogic relationships with others, that some independent posts may be designed to respond to course contents or discussion prompts and therefore create valuable user-content interactions, rather than aiming to engage in user-user interactions.

After examining the potential start of a dialogic conversation, in Chapter 8 the investigation turns to its development by examining the replies. A keyword analysis of replies shows how users engage in stance-taking as they agree or disagree with each other. The finding is extended by micro-analysis of selected threads to illustrate how users build on each other's contributions to explore their differences and co-construct intersubjectivity. How users deal with their entrenched differences is further investigated in the case of the phrases *agree to disagree/differ*, where users reconcile, disengage or recognize their differences in their conversations. This chapter further shows that users' discourse matters for their interactions with others in online discussions, in this case the shift from disagreement to exploration of differences necessary for a dialogic conversation.

Informed by the keyword analysis of replies, Chapter 9 explores another meaning-making resource besides language – URL-posting, which accounts for the circulation of (mis)information in online spaces. Building on previous studies about information use in online spaces, the micro-analysis of *link wars*, where URLs become a point of contention, sheds light on the tension between users who differ in their use of URLs for evidencing practices in online discussions. Ultimately, this finding attests to the role of discourse practices in online discussions, in this case the use of URLs.

The concluding Chapter 10 draws all the findings together to highlight the role of discourse in initiating, sustaining or hindering user-user interactions in online discussions, while acknowledging

user-content interactions. The insights into the differential effects of discourse practices in online spaces speak to the importance of including language practices in digital literacies education to inform users of their agency in establishing social relationships with others in the face of prompt-focused posting, and engaging in dialogic conversations for online deliberation, especially when they hold opposing stances. The chapter outlines empirical, theoretical, methodological and practical contributions of this thesis, and ends with identifying its limitations and future work.

Chapter 2

Research Setting: MOOC online discussions

2.1 Introduction

This chapter continues the introduction of this thesis by detailing the site for the inquiry of online discourse – MOOC – and relevant research. Online discussions take place in various online spaces, so it is impossible to sample every single space. In this thesis, I have chosen online discussions in public educational spaces (MOOCs) as the research setting for investigating online discourse. Given the popularity of MOOCs, users' discourse in its online discussions warrants investigation. Although discourse in online learning might be different from other online spaces given its academic nature (Collins, 2019; Lapadat, 2002), assuming that online discussions in educational spaces are less likely to be subjected to trolling or uncivilized conversations (Littleton & Whitelock, 2005), the findings in this thesis may provide insights into best discourse practices for users' interactions in online discussions in general.

The online discussions in MOOCs have been well researched in the field of educational technology for their implications for learning and technological design (Almatrafi & Johri, 2019; Gasevic, Kovanović, Joksimović, & Siemens, 2014). However, to the best of my knowledge, there is no detailed study on users' discourse practices in initiating and engaging in conversations, especially from the point of view that language does things; that is enact identity, establish relationships and construe meaning. Therefore, this thesis also fills this gap in the educational literature of MOOCs by highlighting the role language plays in users' interactions.

In the following, MOOC and FutureLearn, the MOOC platform examined in this thesis, are introduced first. In relation to the methodological objective set up in Chapter 1, I will then review

MOOC research on users' discourse to show how this thesis will bring another dimension to the field of education, although the main focus of the thesis is online discourse rather than online learning.

2.2 MOOC

A massive open online course (MOOC), as its name suggests, is a course delivered online often without any prerequisite such that anyone who has internet access can join the course. Therefore, a course could easily reach a massive number of internet users based around the world. The same course is also often rerun several times or remains open for access anytime. These courses are typically pitched at university-level and for professional development (Poquet, Dowell, Brooks, & Dawson, 2018). Major MOOC platform providers include Coursera, edX, iversity, and FutureLearn. Depending on their business models, different features are available for free and premium access. It is typically free for users to access learning materials and activities in MOOCs, such as readings, video lectures, exercises and online discussions, while the provision of certificates and assessments is often subjected to fee payment. Given its free and open access nature, not many users complete the course fully (R. Ferguson & Clow, 2015; Hew & Cheung, 2014).

One quintessential feature of MOOCs is the online discussion that allows interactions among users, that has been hailed as beneficial for socio-constructive learning (R. Ferguson & Sharples, 2014). Socio-constructivism was first developed by Vygotsky (1978) who argued that humans co-construct meaning, knowledge and the social world through social interactions with others within a culture. Language is the means of such co-construction. For a similar reason, online discussions have long been utilized as a learning activity in both distance learning and on-campus courses (Lapadat, 2002; Yates, 1996). However, as will be seen in section 2.4, the research on MOOCs mainly analyse user's language to infer thinking and learning, rather than analysing language discursively to understand how users engage in conversations or the socio-constructive process in the conversations.

The online discussion allows the large number of users in MOOCs, who come from different backgrounds and could not have face-to-face interactions, to exchange information, experience and ideas with each other. Their contributions in the online discussion may further enrich the learning experience. This is supported by surveys of users' experiences that found that peer interaction is one of the five factors¹ underlying the popularity of a MOOC (Hew, 2016), and more than half of the users like reading others' contributions in the online discussions of MOOCs (Sharples & Ferguson, 2019). Furthermore, it has been found that users who make many comments and receive responses in the discussion tend to complete the course, attesting to the importance of online discussions in MOOC learning (Alario-Hoyos, Muñoz-Merino, Pérez-Sanagustín, Delgado Kloos, & Parada G., 2016; Coetzee et al., 2014; Swinnerton, Hotchkiss, & Morris, 2017; D. Yang, Wen, Howley, Kraut, & Rose, 2015). Therefore, online discussions are considered as important features in MOOCs and have attracted attention among MOOC and education researchers (Gasevic et al., 2014). However, similar to what has been found in other online discussions, researchers and some users have also voiced concern that user-user interactions in the MOOC online discussions are shallow, fragmented and not critical, which can be traced to the lack of replies and to short-lived interactions (Gillani & Eynon, 2014; Hew & Cheung, 2014; Kellogg et al., 2014; Poquet et al., 2018; Wise et al., 2016).

2.3 FutureLearn

Among all the MOOC platforms, this thesis chooses the FutureLearn platform to investigate online discussions because it particularly emphasises social learning. The FutureLearn platform (www.futurelearn.com) is designed based on the Conversational Framework (Laurillard, 2001, 2012). The Conversational Framework operationalizes learning as an iterative process between reflecting within oneself and conversing with others, while interacting with the outside world, which could be

¹ The other four factors are problem-centric learning, instructor accessibility and passion, active learning and helpful course resources.

the learning materials, abstract concepts, previous experiences or existing knowledge. In FutureLearn, the conversation within oneself and with others occur in the online discussion on the platform which is implemented as “discussion in context” (R. Ferguson & Sharples, 2014), as explained in the following.

In FutureLearn, each course consists of materials for two to eight weeks of learning. Within each week, there are course steps, each of which is like a web page and typically contain one main learning object, which can be video, audio, article, discussion prompt, quiz, test, or exercise. The step can be understood as one unit of learning. In each step, except quiz and exercise, there is a commenting space right below the course content. Figure 2.1 is a screenshot of a step in FutureLearn, which goes on to . also shows part of the commenting space where the online discussion takes place. This is the step from which the three posts at the start of Chapter 1 are drawn. Users are usually encouraged to share their experience, contribute their reflection, discuss issues raised in the course steps, and interact with others in the commenting space (e.g., “There are plenty of opportunities to debate with other learners. You’ll be able to make comments at any point in the course [.....]”² You’ll also notice discussion points, which offer a more structured dialogue with your fellow learners on key topics. Please join in!”³).

² The [.....] is to replace texts quoted from the course or users’ comments that are omitted from presentation in the thesis.

³ <https://www.futurelearn.com/courses/inequalities-in-personal-finance/1/steps/29909>

Figure 2.1 One course step that contains a cartoon and an article.

4.1

YOU'VE COMPLETED 1 STEP IN WEEK 4



© Andy Davey - First published in The Sun, 2010

Pension, housing and wider inequalities

118 comments

So far, you have explored different explanations and solutions to the problem of inequality, in particular focusing on access to a decent pension and to a decent home.

In Week 2, you learned that the problem of creating a sustainable pension system as populations are ageing could be tackled in different ways, although some 'solutions' might be more desirable than others. Examples are:

- increasing levels of individual savings or contributions to a pension fund
- increasing taxation to fund pension systems
- reducing pension replacement rates
- working longer
- increasing the number of migrant workers
- increasing fertility

Note. Due to space constraint, not all the content of the article is shown.

Source: <https://www.futurelearn.com/courses/inequalities-in-personal-finance/1/steps/29219>

Figure 2.2 The commenting space below the course content in each step.

Over the last three weeks you have learned that inequalities in access to housing and in access to a decent pension in old age are influenced by and have an impact on wider socio-economic inequalities. This week returns to these wider inequalities and explores the relationships between different levels of intervention, and how these have evolved over time and could evolve in the future.

Kick-start this week by reminding yourself what is at stake.

© The Open University

118 comments

Mark as complete

<
SUMMARY OF WEEK 3 ARTICLE

SOCIAL PROTESTS AGAINST INEQUALITY ARTICLE
>

COMMENTS

Shi Min Chua

Add a comment... (plain text only, links will be auto-linked)

0/1200

Show: All comments Sort by: Newest

LT

Oh dear! - This course does seem to be turning into a one-sided left wing love-fest!!!! OU, please be more objective and complete.. 3 comments by Stiglitz alone surely cannot sum up all angles on this topic? Remember that there are a lot of people here who obviously do not have much knowledge of economics and so take what they read here at face value. Dangerous if all they get is a few quotes from Stiglitz and Wilkinson!!!! (edited)

Like 1 Reply

VX

The right wing press put the other points of view on a daily basis!! The Ou is just making sure we know there are other ways to analyse the story!

Like 2

Follow

03 APR

I so agree with your commentd. I too expected less bias from ou..i expected for them to give both sides of the equation. I do not read the press as feel each paper wants to indoctrinate their values and although I am one side of the political spectrum I do not agree with all of that party's views.

I was lucky enough to have two teachers for economics each with a different political bent!

there are newspapera tgat covet both sides of the political spectrum and there are those that sit ob the fence. In addition to this there is one I can think of that will knock the government of the day (edited)

Like 1

Add a comment...

MJ

Deregulation in both in the labour market and wage agreement arena have proven to be nothing short of disastrous for the UK. The explosion in more part time work and zero hours contracts has given rise to a far less equal society than we have had in many, many years. The distribution and inequality in income will only get worse unless more full time, well paid jobs are created - whether that be through the amalgamation of part time jobs or the outlawing of zero hours contract.

Like 3 Reply

Creating a new post

Adding reply after other replies or underneath a new post.

Note. Also shown are the different ways that users can add a comment. After the data collection, the platform changes the "Add a comment..." after other replies to "Add a reply..."

24 | Page

In the commenting space, as can be seen in , a user could add a comment separately from others' comments, which I describe as creating a new *post* to differentiate it from *reply*. A reply is added underneath a post. Although *comment* is the terminology used in FutureLearn, I will use it when referring to both *post* and *reply*, and keep the differentiation between *post* and *reply*. Each new post has the potential to elicit replies from other users and initiate a thread of discussion. Those that elicit replies are termed *initiating posts*, and those that do not *independent posts*. An initiating post and the replies underneath it form a *thread*. In Futurelearn, there is no threaded structure among replies underneath a post, and the replies are ordered by the time of posting. Besides adding comments, users can also "like" a comment, or follow other users, and sort the comments based on dates of posting or number of "likes" received. Furthermore, they can filter for their own comments and those contributed by the users they follow. Users will get a notification whenever another user replies to their posts, replies after their reply within a thread, "likes" their comments or follows them.

Under the design of "discussion in context", users commenting in the context of a content is similar to those underneath YouTube videos or news website articles, attesting to the validity of investigating FutureLearn MOOC online discussions as a case of general online discussions. Nonetheless, it should be noted that each step within a FutureLearn MOOC follows a course structure and is visited in the context of a course, whereas YouTube videos or news website articles are individual webpages. Furthermore, most of the time, educators who design the course, academics or post-graduate students with relevant expertise, also join the commenting space as facilitators.


The focus of discussions in the commenting space may be directed by the discussion prompt or the course content in that step, but users are still free to post anything they like. The steps can be roughly categorized to two types: (1) steps that are content-oriented, labelled as video, article, or audio on the platform (Figure 2.1 shows such a step); and (2) steps that are discussion-oriented, labelled as discussion on the platform, and the content consist of mainly discussion prompts, as

shown in Figure 2.3. Some steps may not be as clear-cut, such that there are also discussion prompts in content-oriented steps, as shown in where there is a video and discussion prompts.

Figure 2.3 One course step that contains mainly discussion prompts.

4.27

YOU'VE COMPLETED 1 STEP IN WEEK 4



© monkeybusinessimages (via iStockPhoto.com)

A change is gonna come – your opinion

169 comments

So, what do you think about the different alternative models we have discussed in the final week of the course?

Write a paragraph explaining your opinions and post it in the discussion below.

Use these questions to help:

- Which (if any) aspects of these alternatives seem feasible or desirable?
- To what extent do you think they would be effective in reducing inequalities between and within generations?
- What ideas do you have about how to address inequalities?

Our facilitators are unable to respond to all comments, so please 'like' the comments you want to bring to their attention.

In the next step, review your learning in the end-of-course test.

© The Open University

Source: <https://www.futurelearn.com/courses/inequalities-in-personal-finance/1/steps/29335>

Figure 2.4 One course step that contains a video and discussion prompts.

2.3

5 MORE STEPS TO GO

[View transcript](#)[Download video: standard or HD](#)

Make, buy or outsource?

186 comments

So do we make, buy or outsource? And if we outsource, what can we outsource and what should we never outsource?

In this video, Douglas guides you through the benefits and problems associated with making, buying, outsourcing and global sourcing.

After watching this video, what are your thoughts?

A copy of the mind map used in this video is available to download in PDF format from a link at the bottom of this page.

© University of Southampton 2015

Source: <https://www.futurelearn.com/courses/contract-management/4/steps/115154>

According to the design of the FutureLearn (R. Ferguson & Sharples, 2014), the “discussion in context” is designed to trigger two levels of conversations: (1) conversation with oneself, which could be considered as a self-reflection about the learning materials; (2) conversation with others, which is an explicit dialogue with others. These two levels correspond to the user-content interactions and user-user interactions in other web pages such as the commenting space in Youtube, news websites and Facebook (Frobenius & Harper, 2015; Herring, 1999; Ksiazek & Lessard, 2016; Ziegele, Breiner, & Quiring, 2014), where users could either respond to the content of the page or to other users’ comments.

2.4 Research on MOOC online discussions

There have been numerous studies on various aspects of MOOCs and their online discussions (see Almatrafi & Johri, 2019; Gasevic et al., 2014 for a review). For the purposes of this thesis, it is neither possible nor necessary to review and consider every aspect of MOOC online discussions. Rather than examining MOOCs or educational implications, I use MOOC online discussions as an online space for examining online discourse. Therefore, only MOOC research that examines users’ textual contributions, that is comments, are reviewed here. However, I argue that the focus of this thesis will also bring a new methodological perspective to the research field of MOOC online discussions, and provide insights for users who join the online discussions in MOOCs.

2.4.1 Coding and counting

To the best of my knowledge and based on the systematic review by Almatrafi & Johri (2019), most studies on MOOCs use a coding and counting paradigm of content analysis to analyse users’ comments in the online discussion, either manually or automatically. This paradigm is also widely used in online discussions in other learning settings, such as distance learning or university courses (see Loncar et al., 2014 for a review). To some extent, these studies categorize or code what comment is or represents – for example, whether a comment is on-topic or off-topic (Wise et al.,

2016), whether a comment shows that a learner engages in higher order thinking or pays attention (Wang et al., 2016) or whether a comment is positive or negative (Wen, Yang, & Rosé, 2014) – rather than *how* a comment is written. The comments typically are analysed individually, rather than discursively to understand how a conversation is co-constructed by learners. In another MOOC study (Kellogg et al., 2014), instead of individual comments, each thread is coded into one of the five levels, from the lowest levels of sharing to the highest levels of metacognition to examine the extent of knowledge co-construction of the whole threads.

In these studies, users' textual contributions are reduced to codes for counting purposes such that the online discussions can be quantified and (cor)related to other variables statistically, such as learners' learning outcome, participation pattern, or course design. The coding also forms the basis for automatic recommendation of quality comments to users, monitoring of the discussion space, and prediction of learning performance (Almatrafi & Johri, 2019; Wise et al., 2016). These studies are useful for evaluating the quality of the online discussions, and may inform educators or designers for further improvements of their MOOCs.

However, equally important is the dialogic and socio-constructive processes that are realized by users' actual language and discourse practices. Several researchers, including educational researchers, have argued that users need to acquire digital language competency to engage in online discussions, especially for negotiation (Laflen & Ftiozenza, 2012; Landqvist, 2016; Littleton & Whitelock, 2005; Paulus, 2006). For example, in a small-scale online learning discussion, sharing of experience can be a way to reach agreement and affiliation in co-construction, or can be rejected by others as authoritative, depending on how users orient it in the on-going discussion (Kääntä & Lehtinen, 2016). This finding, although not based on MOOCs, highlights the importance of investigating the comments discursively within the threads as well as the role language plays in the co-construction process, in this case *how* the experience is shared, not just *what* the comment is about – sharing experience. It also attests to the proposal that users' discourse practices in online spaces are one area of digital literacies that warrant attention (Jones & Hafner, 2012).

Arguably, in these MOOC studies, users' language is the basis for coding. However, the language in the comments is used to infer thinking and learning by assuming that texts are static and exact representation of users' inner cognition (Wegerif & Mercer, 1997; Wise & Paulus, 2016). Based on this assumption, the actual language practices users employ are not the focus of these studies. Thus, there is no revelation of the social-constructive process or the dialogic nature of users' discourse in these studies, that is how a conversation unfolds from one reply to another within a thread. To fill this gap in MOOC research, this thesis instead takes the position that language does things and follows the assumption of discourse analysis that "Text can only tell us what people do (and not what they really think or feel)" (Herring, 2004, p. 359). More specifically, this thesis aims to understand how users use language to initiate and engage in dialogic conversations in online spaces.

2.4.2 Corpus linguistic approach to MOOCs

Admittedly, the main aim of MOOC research is the educational implications, so it has a different aim from this thesis, thus a different choice of methodology and position. It is also possible that the large number of users' comments in MOOCs renders a detailed analysis of discourse practices infeasible, whereas the coding and counting paradigm effectively reduces the language data to manageable variables. To harness both the big data available from MOOC online discussions and the rich language data to explore discourse practices employed by users, this thesis will introduce a mixed methodology, corpus linguistics, to MOOC research.

As far as I am aware, there is only one corpus linguistic study conducted on one MOOC. Collins (2019), who was also one of the educators of the FutureLearn MOOC "How to read your boss", employs established corpus linguistic methods, keyword analysis, concordance and collocation analysis (see Chapter 4 for explanation on these methods), to investigate users' use of a technical term taught in the course, *face*, in their comments to explore evidence of their learning. In his corpus, *face* is found to be a keyword that is used significantly more often when compared with British National Corpus of written English (BNC, Leech, Rayson, & Wilson, 2001). However, it is used

mainly when users respond to the discussion prompt, “Is ‘personal face’ or ‘social identity face’ more appropriate for your workplace?” in a course step. This is evidenced by the collocates of *face*, that is words co-occurring with it, *social identity*, which also appears in the discussion prompt. Collins (2019) finds that this technical term is seldom used in other contexts and other course steps, suggesting that users might not have acquired the concept of the term to generalize it to other contexts. However, he does not investigate other keywords found, such as *I, my, boss, work, think*, probably because he is only interested in understanding users’ use of the concept of *face*.

Of relevance to the aims of this thesis of investigating online discourse in user-user interactions, Collins (2019) also finds that users’ discourse surrounding the keyword *face* is dialogic in nature. This is evidenced by concordance reading, that is examining the co-text⁴ where the keyword *face* occurs, which indicates users explicitly invite others “Would somebody explain the difference between Personal Face and Social Face?” (p. 142) or express their uncertainty “I don’t feel I fully understand the difference between personal and social face” (p. 142). However, he does not further explore this aspect of users’ discourse.

Besides investigating the corpus as a whole, Collins (2019) also investigates keywords in posts and replies to explore interactivity in the online discussions, by comparing each type of comments to BNC. Keywords found in the posts include *identity, boss, personal, I, am* whereas keywords found in the replies include *hi, i_agree, you, I, am*. However, he does not go into details to explore linguistic features and discourse practices realized by these keywords. Furthermore, he does not differentiate between initiating posts and independent posts. Collins (2019) himself also suggests more in-depth discourse analysis that draws on conversation analysis, is needed to understand users’ conversations in this online space.

Collins (2019) successfully shows that, besides the often-used coding and counting paradigm, a corpus linguistic approach can reveal textual evidence of users’ learning and interactions

⁴ The words surrounding a particular word.

in MOOCs. However, this study is limited in scope as only one MOOC is investigated and focuses on only one concept taught in the MOOC. This study definitely shows that educators can use the corpus methods to examine learners' learning in the MOOCs they teach. This study also shows a preliminary finding that discourse in posts and replies are different, although how users employ the discourse to engage in conversations remains unexplored.

2.5 Conclusion

This thesis will build on Collins' (2019) corpus study in this respect of language practices in the posts and replies to further examine the dialogic nature of online discourse, and at the same time fill the gaps of mainstream MOOC research that have not examined users' discourse practices. More importantly, this thesis will further dissect posts into initiating posts and independent posts, replies into first contributions and subsequent contributions, as will be discussed in the next two chapters.

Furthermore, unlike Collins' corpus analysis or other MOOC studies which only examine one MOOC, or multiple presentations of the same MOOC (Poquet et al., 2018; Wang et al., 2016; Wise et al., 2016), this thesis investigates 12 FutureLearn MOOCs of different disciplines in an attempt to achieve a balanced view of online discussions and online discourse. In Chapter 4 and 5, I will show how an 11-million-word corpus of comments from the 12 MOOCs was compiled and interrogated both quantitatively and qualitatively to examine online discourse. Before going into the methodology, first in Chapter 3, I will review literature on user-user interactions, online discourse and URL-posting, and theoretical concepts directing the analysis of users' discourse practices.

Chapter 3

Research into online discussions and the dialogic nature of online discourse

3.1 Introduction

This chapter reviews research on online discussions and two theoretical concepts directing the inquiry of the dialogic nature of online discourse. Literature of corpus studies on online discourse will be reviewed in Chapter 4.

This chapter starts with a literature review to highlight the extent of lack of replies and short-lived interactions that may compromise information exchange, socialization and deliberation in online discussions. Although research has begun to investigate how users can frame their comments online to establish conversations with others, these studies only investigate linguistic features that predict the chance of receiving replies. In-depth analysis of discourse practices on how users could invite replies and establish conversations is still largely missing. This thesis thus aims at extending these studies by systematically examining the linguistic features and discourse practices in posts that receive replies and those that do not.

Previous research also examines user-user interactions in long and sustained threads in online discussions for their thread evolution, polylogue and negotiation processes. The findings on long threads speak to users' agency in navigating the complicated nature of online discussions, while enacting their identity and relationships with others. This thesis will further investigate users' discourse practices within threads by systematically examining the linguistic features and discourse practices in replies. Overall, the first half of the literature review points to the empirical need to investigate the dialogic nature of online discourse that are facilitative of conversations in online spaces.

Besides language, there are other semiotic resources at users' disposal in online spaces, for example emoticons, GIFs and hyperlinking. This thesis thus also examines URL-posting in online discussions, which is afforded by most MOOC platforms (including FutureLearn). As we shall see in the literature review of URL-posting in online spaces, most of these studies only focus on *what* types of links are posted with limited analysis of the discursive context in which the URLs are posted. This thesis will expand these studies by examining *how* users employ and respond to URLs in their conversations.

In the rest of the chapter, I will introduce two theoretical concepts – dialogic space and intersubjectivity – which direct the analysis of user's discourse practices in online discussions to understand the dialogic nature of online discourse, thereby addressing the research questions:

RQ1: What are the differences in the linguistic features and discourse practices that regularly occur in

- initiating posts that receive replies and start a discussion thread,
- independent posts that do not receive replies,
- replies, especially those in sustained discussions?

RQ2: How do these discourse practices initiate, sustain or hinder dialogic conversations in online discussions?

RQ3: How does URL-posting initiate, sustain or hinder dialogic conversations in online discussions?

In the context of these two theoretical concepts, relevant discourse practices, the value of sustained threads and disagreement in online discussions are also discussed to highlight the importance of dialogic discourse in online spaces.

3.2 User-user interactions in online discussions

As argued in Chapter 1, the dialogic nature of online discourse is most likely found in user-user interactions in online discussions. To engage with others in online discussions, users can create new posts to invite others' replies, for example to seek advice, information or shared experience. Users can also reply to others' posts or threads to engage in a conversation, whether to provide advice, information, socialize or deliberate on issues. The extent of user-user interactions in an online space can be gauged from the distribution of posts that receive replies, number of replies and length of threads (Ksiazek & Lessard, 2016; Lewis, 2005; Ziegele et al., 2014). Another possible indicator of the extent of, or lack of, user-user interactions are the number of users who contribute only once in a single online space, for example the commenting space below a news story or a course step in FutureLearn. This is based on the assumption that if a user just posts once and never comes back, it is hardly possible to have a conversation with other users, not to mention to engage in deliberation (Freelon, 2015; Herring, 2004; Ruiz et al., 2011).

Previous research that present these indicators, however, have shown that the extent of user-user interactions might not be sufficient for the functions of socialization and deliberation in online discussions, as shown in Table 3.1. Firstly, in some online discussions (Ksiazek & Lessard, 2016; Sunar, Abdullah, White, & Davis, 2015; Tubman, Oztok, & Benachour, 2016), especially those with commenting space alongside content, only about one fifth of the postings are replies, suggesting that very few users engage in explicit interactions with others via replying. Secondly, in some online discussions (Beth et al., 2015; Burke et al., 2007; Marcoccia, 2004; Meyer et al., 2019), about half of the new posts do not receive any replies, suggesting that some users may fail to elicit responses from others to start a conversation. This is not the case in all studies: in two studies (Arguello et al., 2006; Cui, Jin, & Wise, 2017), not as many new posts were left without any replies. However, it is worth noting that Arguello et al. (2006) only sample the first new post posted by each user, so it is possible that this sampling underestimates the number of new posts that do not receive replies, assuming that some users post more than one new post to get replies (Herring, 1999). In Cui

et al.'s (2017) study on a MOOC, interestingly, although 86% of new posts receive replies, 61% of them receive just one or two replies, suggesting only short-lived conversations are generated.

Thirdly, besides Cui et al. (2007), a few studies (Beth et al., 2015; Marcoccia, 2004; Napoles et al., 2017; Tubman et al., 2016) have reported that conversations in online discussions are generally fewer than five replies. This raises the question of how negotiation processes for deliberation could occur within such limited turn-taking (Poquet et al., 2018), assuming that each turn corresponds to the initiating post and every reply within a thread. Fourthly, some users only contribute once in the online discussion space (Bou-Franch & Garcés-Conejos Blitvich, 2014; Ruiz et al., 2011). This might indicate that the users post in response to the content, i.e., forming user-content interactions, rather than engaging in a continued discussion.

Table 3.1 Selected studies that have reported quantitative information regarding user-user interactions

Study ¹	Focus of the study	Online Space	Indicators of user-user Interactions
Marcoccia, 2004	Online polylogues	Usenet newsgroup	50% of new posts do not receive any reply.
Arguello et al., 2006	Factors affecting users' willingness to respond to others' posts	Usenet newsgroup	27% of new posts do not receive any reply; no information for length of threads.
Burke et al., 2007	Rhetorical strategies to elicit replies	Usenet newsgroup	43% of new posts do not receive any reply.
Ruiz et al., 2011	Democratic qualities of online news discussions	News website commenting space	85% users contribute only once.
Bou-Franch & Garcés-Conejos Blitvich, 2014	Conflict management	YouTube commenting space	69% users contribute only once.
Beth et al., 2015	Learners' sense of responsibility to contribute	Discussion forum in virtual learning platform	44% of new posts do not receive any reply; 33% of new posts receive only one reply.
Sunar et al., 2015	Recurrent interactions among learners	FutureLearn commenting space	17% of all comments are replies.
Ksiazek & Lessard, 2016	User engagement in online news	YouTube news video commenting space	19% of all comments are replies.
Tubman et al., 2016	Sociocultural analysis of learning	FutureLearn commenting space	Of all comments across 10 courses, number of replies ranges from 19% to 49.91%; Only report length of threads in one course, in which 84% contain only one or two replies.
Cui et al., 2017	Content-related interactions	MOOCs on Stanford open-source platform Lagunita	14% of new posts do not receive any reply; 61% of threads only have one or two replies.
Napoles et al., 2017	Discourse of argumentation	Yahoo News comments	Average length of threads is 3.8 comments ² .
Meyer et al., 2019	Attitudes and beliefs about influenza vaccination	News website commenting space	43% of new posts do not receive any reply;

¹Ordered by year.

²The study does not specify if the length of a thread includes the initiating post.

Social network analysis has similarly demonstrated rather short-lived and fragmented user-user interactions in MOOC online discussions. At an individual level, Sunar et al. (2015) find only 1.75% of users have recurrent interactions with other users across a MOOC course period, whereas Poquet et al. (2018) find that even users who comment regularly across the course period, only have an average of four recurrent interactions. At the level of social structure, both Kellogg et al. (2014) and Gillani & Eynon (2014) find that the network in MOOCs is rather dispersed and there are users who are not connected to any other users, i.e., not receiving reply or replying to others. These findings suggest that users may not have any conversation with others or subsequent conversations with the same users in MOOCs, probably because the large number of users and comments render the chance of them replying to each other less likely (Himmelboim, 2008) or because a large number of them drop out of the MOOC (R. Ferguson & Clow, 2015).

These quantitative data signify the tension between user-content and user-user interactions. The large number of new posts instead of replies in some online discussions can be an indication of *prompt-focused* posting such that users mainly respond to content or prompts on the page rather than other users (Herring, 2013). It might also be an indication that users are more interested in expressing their view independent of others' views (Bou-Franch, 2012; Cavanagh, 2007). In these cases, users might be more likely to engage in user-content interactions, which can be a sign of successful engagement by those hosting the discussion space (Preece & Maloney-Krichmar, 2005).

However, the prompt-focused posting of user-content interactions can compromise user-user interactions in online discussions, and reduce the chance for dialogic conversations to occur (Ksiazek & Lessard, 2016; Preece & Maloney-Krichmar, 2005). Lack of responses to a user who is seeking information and support, or socializing can be off-putting (Crook et al., 2016; Hew & Cheung, 2014), and may result in newcomers giving up posting (Joyce & Kraut, 2006). Compared to user-content interactions, online users prefer user-user interactions and consider it as more informative (Pace, Buzzanca, & Fratocchi, 2016). Specifically in a MOOC discussion, D. Yang et al. (2015) found that users who receive replies that address their confusion are less likely to drop out. Some one-time

contributors could possibly be those who do not receive replies to their first post and so stop posting (Herring, 1999).

Furthermore, as shown in some of the above studies, although some posts receive replies, they are often short-lived threads and lack recurrent interactions to be conducive for negotiation to take place. This phenomenon unfortunately goes against the hopes of some users and researchers for online deliberation and co-constructive processes (Delahunty, 2018; Poquet et al., 2018; Springer et al., 2015). Continuous conversations in sustained threads are needed for negotiation and intersubjective processes to occur and will be discussed further in section 3.4. In short, conversations among users are considered as one important indicator of the success of an online community (Burke et al., 2007; Preece & Maloney-Krichmar, 2005).

3.3 Evolution of threads and polylogal nature of threads

Despite the often-occurred short-lived threads, long and sustained threads that demonstrate continuous user-user interactions do take place in online spaces. These threads have been examined in two aspects: 1) evolution and polylogal nature of threads, where polylogue refers to conversation among multiple interlocutors within the same thread; 2) how users respond to each other within the threads (Herring, 2004; Jaworska, 2018; Ziegler, Paulus, & Woodside, 2014). The first aspect can be considered as more about the structure of a thread, whereas the second aspect is regarding discourse practices. These two aspects are generally investigated with micro-analysis such that the text and language of the initiating post and all the replies within a thread are analysed in relation to each other (Giles, Stommel, Paulus, Lester, & Reed, 2014).

In this section, the evolution and polylogal nature of threads are discussed to give an overview on user-user interactions in threads, and provide evidence for users' agency in achieving interactions amid the "messy" interactions within a thread. More importantly, the dynamic in the threads attest to the need of micro-analysis of threads, instead of investigating posts or replies

separately. Discourse practices regarding how users respond to each other within a thread will be reviewed in section 3.5.

In a long thread, the topic originally started by an initiating post typically evolves. Firstly, a thread could develop or drift into a new topic, either related or not related to the original topic, such as hostility, humour, metapragmatic discussion, i.e., discussion about the discussion (Herring, 1999; Lambiase, 2010; Potter, 2008; Tanskanen, 2007), or political talks found in non-political forums (T. Graham, Jackson, & Wright, 2016). Secondly, the topic of the thread may stay relevant, yet branches out to other relevant sub-topics (Herring, 1999; Lewis, 2005; Thomas, 2002). These evolution patterns can happen in the same thread, and interleave with each other such that the adjacency pair of turns, i.e., consecutive replies, are not necessarily on the same topic. Users and those hosting the discussion may not share similar interpretation and preference towards the evolution of a thread. For example, a thread that starts from a research-based topic and drifts to anecdotes and personal experience can be deemed as relevant by some but a decay to non-relevance by others (Potter, 2008).

Topic evolution is also associated with the polylogal nature of threads. This can be due to the asynchronous nature of online discussions, such that any user can join in the same thread at any time. Users are not bound to only reply to the latest reply in a thread, but can also reply to the initiating post, a specific reply or several replies in the thread in one go. The polylogues can arise with the evolution of the thread. Within a thread, each topic or sub-topic can take place in a one-to-one dialogue between two users, or in a many-to-many polylogue among multiple users (Bou-Franch & Garcés-Conejos Blitvich, 2014; Marcozzia, 2004). The replies forming these dialogues and polylogues may interleave if they are not posted consecutively in the thread (Herring, 2013). Besides these two patterns, polylogues within a thread can sometimes be many-to-one or one-to-many conversations. For example, multiple users can collectively respond to the initiator or to a particular user who holds a strong opposing stance (Herring, 1999; Lewiński, 2010).

As with topic evolution, polylogues in a thread can be changing constantly. For example, another user can join in an on-going one-to-one dialogue and change the dynamic to polylogue, or a polylogue among multiple users can eventually drift to a dialogue between only two users (Lewis, 2005). There are also times that some users' replies are not picked up among all the replies within a thread, and are thus seen to be excluded from the conversation (Herring, 1999), whereas there are also users who are in dialogue or being addressed by others but do not come back to the thread to continue their conversations (Bou-Franch, 2012; Herring, 1999; Kleinke, 2010; Lewinski, 2010; Maroccia, 2004). Fortunately, users other than the addressed users can respond given the polylogal nature of online discussions (Bou-Franch & Garcés-Conejos Blitvich, 2014).

The topic evolution and polylogue in a thread paint a rather complex and messy picture of interactions in online discussions. This seeming messiness can be caused by the threading system that does not allow further threading to indicate relationships between replies (Bou-Franch, 2012; Herring, 1999). However, the flourishing of topics and polylogue under this technological design suggests users' agency in achieving interactions with each other. Within polylogal threads, users are found to have adapted their discourse practices and used linguistic features to maintain coherence in the threads, such as vocatives, turn-management devices, quoting, repeating, backchanneling, referencing to previous messages, marking digressions and meta-language (Baym, 1996; Bou-franch, 2012; Frobenius & Harper, 2015; Herring, 1999; Lapadat, 2007; Lewis, 2005; Tanskanen, 2007; Thomas, 2002).

The multiplicity of topic and polylogue can actually be appealing to users in online discussions (Benwell & Stokoe, 2006; Faraj, Jarvenpaa, & Majchrzak, 2011; Frobenius & Harper, 2015; Herring, 1999; Wegerif, 2010). Users can engage in multiple topics or sub-topics in the same thread or different threads at the same time, thus not being restrained by only one dialogue, as in typical spoken conversations. They can selectively respond to any of the replies within the thread rather than only the latest one (Bou-Franch, 2012; Lewiński, 2013; Lewis, 2005; Lorenzo-Dus, Garcés-Conejos Blitvich, & Bou-Franch, 2011; Maroccia, 2004). The drift to humour or topics of interests,

and creative language play in online discussions also suggests users' positive experience, rather than frustration with the complexity (Herring, 1999). Furthermore, some researchers (Mercer, 2004; Potter, 2008; Ugoretz, 2005; Wright, 2012) also contend that a digression to other topics can be an exploration that leads to productive discussions of different perspectives or preferred topics. Users might be 'forced' to read various sub-topics when participating in the thread, with the potential of broadening their viewpoints and avoiding echo chambers (Dron, 2014; Walter et al., 2018; Wegerif, 2010). Therefore, multiplicity of topic and polylogue are not only appealing to users but an opportunity for them to be in touch with different views and topics. This view contrasts with other researchers (Coetzee et al., 2014; Wise et al., 2016) who argue for technological interventions that automatically filter or organize comments. Instead, the literature on topic evolution and polylogue, as well as discourse practices to be reviewed later in this chapter, suggests users' agency in the online discussions. This is the perspective taken in this thesis, while acknowledging the technological affordances and constraints of online spaces.

3.4 Linguistic features and content characteristics of posts

To investigate how users could invite replies to their new posts and initiate a conversation with others, thus, to be involved in user-user interactions in online discussions, several studies have examined the linguistic features or content characteristics of posts that receive replies. These studies mainly used a coding and counting paradigm to operationalize language data into variables to be subjected to a statistical analysis to predict the chance of receiving replies. However, detailed analysis of discourse practices is missing for understanding the dialogic nature of online discourse.

It is found that controversial content (Roederkerk & Pauwels, 2016), imported content from news websites by URLs (Himmelboim, Gleave, & Smith, 2009), correct and incorrect ideas (Chen, Lo, & Hu, 2020), and relevance to the community (Arguello et al., 2006) increase the chance of receiving replies. In terms of language characteristics, which is the focus of this thesis, two studies (Arguello et

al., 2006; Crook et al., 2016) detect linguistic features that are likely to increase the chance of receiving replies with a word-based automatic analyser. Both studies similarly find that first-person pronouns, expressions of negative emotions (e.g., *cry, hate, enemy, nervous*) and of cognition mechanism (e.g., *cause, consider, think, know, maybe, always*) increase the chance of receiving replies. But these two studies differ in other linguistic features found: Crook et al. (2016) find present and past tense verbs, whereas Arguello et al. (2006) find third-person pronouns and expression of positive emotion (e.g., *happy, pretty, good*) increase the chance of receiving replies. This difference could be because the online discussions examined are also different: Arguello et al. (2006) examine eight Usenet newsgroups covering health, sport and political topics whereas Crook et al. (2016) examine an online young adult cancer survivorship support community. However, more importantly, both studies stop at the linguistic features found, without examining the discourse and context where these features are used, such that there is no way to interpret this difference and how these features increase the chance of receiving replies.

In another study on 99 Usenet groups that contain health, technical, hobby, and political discussions, Burke et al. (2007) zone into discourse strategies where first-person pronouns are used. They find that making requests with self-introductions that signal one's commitment and affiliation to the group is more likely to receive replies than self-introduction without such signals. This study suggests that the same linguistic features could be used differently discursively and may have different effects in their probability of inviting replies.

Besides linguistic features, Chen & Chiu (2008) and Chen et al. (2020) have developed a statistical discourse analysis to predict whether linguistic behaviours as shown in a message, which they do not really specify if it is a post, will receive replies. The linguistic behaviours investigated include agreement vs disagreement towards previous messages, or positive vs negative affective expression, question, command or justification. The linguistic behaviours are coded based on users' textual contributions and subjected to statistical analysis, similar to the coding and counting paradigm. In an online academic discussion on maths, Chen et al. (2020) find no evidence of emotion

expression, neither positive nor negative, increases the chance of receiving replies. They also find that disagreement, rather than agreement, and question rather than command increase the chance of receiving replies. Although this study is useful in informing users on how to write their comments online, it is still limited in revealing *how*, for example, a question is asked, or *how* to raise a disagreement to open up rather than close down a dialogic space, as will be discussed in section 3.6.

This thesis builds on the above studies by not only examining linguistic features, but also deriving discourse practices likely to invite replies from the linguistic features found in initiating posts and independent posts based on established theoretical concepts – dialogic space and intersubjectivity (section 3.7). This will not only facilitate the interpretation of the actual language practices employed by users, but also provide a more complete picture on how users could initiate conversations in online discussions.

3.5 Discourse practices in replies

A thread or a conversation is not only initiated by posts, but is also developed by replies. Therefore, the linguistic features and discourse practices in users' replies will also be investigated in this thesis to explore the dialogic nature of online discourse. The studies that explicitly focus on users' replies in online spaces are those on users' agreement or disagreement towards others' messages (Baym, 1996; Kleinke, 2010). Baym (1996) examines how users signal their agreement and disagreement, both explicitly or implicitly, to others in Usenet groups. Among the messages that are explicit responses to other messages, that is replies, 33% are identified as expressing (dis)agreement. It should be noted that these replying messages can contain other components, and Baym only examines how a message is positioned as (dis)agreement, and the function of such expressions in the messages.

Baym finds that an explicit expression of agreement can function to refer to previous messages, socially align with others, express gratitude to others who voice one's own view, or

acknowledge others' views. There are also times that users first agree explicitly then raise their opposing views. On the other side of the same coin, there are times users do not express agreement but their elaboration or reasoning are in line with others, possibly to indicate that they are contributing new information in the online spaces, or to move away from a point of conflict. There are also times that replying users seem to be aligned with the previous messages, but downgrade their agreement with contradictory examples or personal experience. All these findings point to the fact that expression of agreement does not necessarily reflect users' thinking, but constitutes different ways of managing their social interactions and conversations. There are also various ways of raising disagreements in the newsgroup. Baym concludes that disagreement seems to be mitigated when users qualify their argument to leave room for other views, or when they provide elaboration and reasoning. Interestingly, addressing other users by name seems to occur less often in disagreement than agreement, possibly to reduce the negative connotation to a person brought about by disagreement.

Baym's (1996) study is valuable for our understanding of how users respond to each other in asynchronous text-based online discussions. Baym shows that users can voice their own view while managing social relationships at the same time, despite holding onto different views. However, Baym acknowledges that she only examines the messages containing components that (dis)agree with previous messages. In other studies, the replies are usually examined in the context of threads or a particular discourse practice, instead of being a focus itself. For example, in the context of examining *experience talk* rather than replies, Kääntä & Lehtinen (2016) find that the reciprocal sharing of experience through first and second stories by users in consecutive replies within a thread is another way of expressing agreement and affiliation with others. Other users' discourse practices in replies will be further discussed later in section 3.6.1 on intersubjectivity and section 3.7 on sustained interactions.

To explore other possible language use in replies, this thesis examines linguistic features and discourse practices in replies based on the keywords found in the replies, rather than restricting

itself to one particular discourse practice. Lastly, it is worth noting that Baym's (1996) analysis of users' responses to others draw upon findings from conversation analysis of spoken conversations (Pomerantz, 1984). Similarly, when examining replies, as well as the threads, this thesis will follow such an approach, as will be explained in Chapter 4.

3.6 URL-posting in online spaces

URL-posting is specific to online spaces given the technological affordance of hyperlinking (Tyrkko, 2010). On one hand, users can easily create a hyperlink or share a link in various online spaces, including social media, blogs and online discussions, and can click on any hyperlink to have direct access to text, pictures, videos and other information external to the immediate communicative context. In other words, hyperlinking is a way of creating intertextuality (see section 3.7) that expand dialogic space from the current space to another (Kiernan, 2018). On the other hand, various websites, including news websites, YouTube, blogs, and websites of private companies or institutions, typically have various options to encourage users to share their pages via social media, private messaging apps or emails.

Specific to the research setting of this thesis, i.e., Futurelearn online discussions, users must fully display the URL address in their comments for hyperlinking, as shown in Figure 3.1. Thus, some users might employ phrases such as "can be found at this link..." to explicitly incorporate the URL address. The platform does not allow hyperlinking embedded in a word, phrase or image, as can be done in blogs or other websites where the hyperlinked source is optional to read and is not explicitly stated in text (Myers, 2009). For example in Wikipedia page on "MOOC", "A massive open online course (MOOC /mu:k/) is an online course..." (Wikipedia, n.d.) where the "online course" is hyperlinked to its page.

Figure 3.1 How hyperlinking to URLs is shown in FutureLearn online discussions.



Note. This post can be found at <https://www.futurelearn.com/courses/palliative/1/comments/16898420>.

The convenience of URL-sharing gives rise to the propagation of misinformation or partisan information online (Burnett & Jaeger, 2008; de Maeyer, 2013; Giglietto, Righetti, Rossi, & Marino, 2020). Although researchers have been trialing algorithms to detect spam links, coordinated link posting and misinformation on social media (e.g., Cao, Caverlee, Lee, Ge, & Chung, 2015; Giglietto et al., 2020), this field of research is not considered in this thesis because this thesis focuses on human's agency in the online world, rather than technological interventions. The literature reviewed here is restricted to those studies that examine the sources of URLs posted by users and how users employ such sources for their discussions online.

Table 3.2 presents selected studies on URL-posting in text-based asynchronous online discussions. It is worth noting that in some studies (Oh, Oh, & Shah, 2008; Savolainen, 2014; Wikgren, 2001), URLs linked to websites are examined along with other information sources referred to by users, including personal opinions or experience, print media or books. Users' reliance on different types of sources seems to vary across different online spaces or topic of discussions. In Yahoo! Answers, users seem to rely more on personal opinions, although online sources are also used (Oh et al., 2008). Similarly, personal opinions written in forum posts or social media that are linked by URLs are also cited frequently in investment forums (Connor, 2013) and medical online forums (Sudau et al., 2014). In contrast, in Usenet health groups, web citations are the most used sources (Wikgren, 2001, 2003). These findings suggest differences in source preferences across

individuals and online spaces, and it has been argued that these differences can create a tension among users (Polletta et al., 2009; Savolainen, 2014; Wikgren, 2001).

Table 3.2 Selected studies that have examined URL-posting in online discussions.

Study	Focus of the study related to URL-posting	Online Spaces	Findings	Points made
Wikgren (2001)	Types of sources used, and how the information is assessed and discussed	Usenet health discussion groups	370 citations or references are found in 160 of the 578 postings (18%). 307 are web citations, 34 are books or printed articles, 29 are references to a person or organization.	Most citations receive no comments on their quality. Comments received are mainly acknowledgements, except a few cases in which other users criticize the sources.
Wikgren (2003)	The citation patterns on a controversial health issue: the beneficial or hazardous use of dietary chromium supplementation in diabetes self-management	Diabetes newsgroup	86 citations to web sources in 14 texts (there is no mention if the text is post or thread).	References to commercial or alternative health websites are met with criticism, whereas abstracts for medical journals are used as “hard currency” to legitimate one’s claims. However, users seldom express the need for reading into the details of the web sources, suggesting that referring to them may only be a rhetorical strategy to win an argument.
Oh et al. (2008)	Sources people prefer to use when they answer questions online	Yahoo! Answers	Among the 101,985 answers, 7,834 answers (7.68%) include source information. 5,391 sources are analysed. Human (56.4%) is the most frequently cited source, followed by online (40%) and offline sources (4%). 71.4% of the internet sources are from commercial/personal websites, while Wikipedia is the most cited source.	Users mainly rely on their own or third party’s prior knowledge, and personal or situational experiences in their answers.
Polletta et al. (2009)	The extent that URL-posting fosters online deliberation	Online deliberation forums	2.8% of 9031 messages contain at least one link. 17% of 549 users post at least one link. 37% of the URLs	URLs are used more often for sharing information and providing alternative ideas for deliberation, rather than for persuading others.

		convened by civic organizations to solicit public input into the rebuilding of the World Trade Center site	are linked to design sites, which is used to encourage others to have a look at the options available. 19% of the links are to magazine or newspaper articles, 15% to advocacy groups. In terms of the discursive context in which URLs are introduced, 33% of the links are used to share information such that they are introduced with little commentary. 43% are used for answering a question or showing an idea. Only 16% are used in an argumentative context. 30% of all the links receive responses from others, most of which appreciation.	Very few users question the credibility of the websites linked to.
Connor (2013)	Types of information cited and valued by participants in online investing forums	Investment discussion forums	Among 1787 posts identified as collaborative and information oriented, 27% cite at least one source. The sources include other posts, books, rating services sources, news sources, and other websites.	The citation of other posts suggests heavy reliance on personal experience. Commercial websites are also heavily relied on but there seems to be no question regarding the trustworthiness of the websites.
Savolainen (2014)	How information sources are used as rhetorical strategies in answering questions about global warming.	Yahoo! Answers	Among 994 answers analysed, 197 references to information sources are found: 35% scientific sources, 30.8% persuasive materials (websites that advocate a particular ideology), 18.6% other sources (e.g., message available on a Q&A site or an opinion of a friend),	Information sources are used most often when users engage in strategies of appealing to authority and reason. Arguments of both warmists' and denialists' nature draw on scientific sources and persuasive materials for their debate on global warming. These sources are used as the factual basis of their argument, and indicative of appeal to authority. The authors argue that there is a constant tension

			8.5% popular scientific sources, 7.1% news. It is worth noting that some of the sources are not necessarily URLs in this study.	between scientific sources and persuasive materials in this online space.
Sudau et al. (2014)	The sources of information used in the discussion on Chronic Cerebrospinal Venous Insufficiency (CCSVI) hypothesis	Online forum of the German Multiple Sclerosis Society	Among 8628 posts identified as relevant to CCSVI, 2829 URLs, mostly linked to social media, are found.	Sources are based on personal experience and opinions, rather than scientific results
Jacobson, Myung, & Johnson (2016)	To what extent linking might intensify partisan political discussion or introduce multiple perspectives	Discussions on the Facebook pages of two partisan cable news organizations	On the Maddow page, 12.86% posts contain 2220 links, on the O'Reilly page, 9.02% posts contain 1222 links. Only 12.2% sources are linked to by both groups, whereas 87.7% sources are linked to by either group. Most links are to mainstream media, followed by blogs and independent media. YouTube is the most linked-to sites (20.8%), followed by Facebook (6.1%). About two-thirds of the links are used to share information or story ideas. Each group posts links more in line with their political ideology.	Users seem to show a preference to a small number of sources. Although some sources are shared on both group pages, each group seems to be only exposed to sources which are line of their ideology, raising the concern of echo chambers.

These studies also show that URLs are posted mainly for information sharing and to a lesser extent, deliberation, based on the discursive context in which the URLs are presented (Polletta et al., 2009; Savolainen, 2013; Wikgren, 2001). Perhaps because the URLs are also perceived by other users as information sharing, most comments containing URLs receive appreciation from others, while users seldom raise questions about the credibility of the URLs (Connor, 2013; Polletta et al., 2009; Wikgren, 2001). In contrast, in those online spaces where users engage in discussions over contentious issues, such as climate change, politics or dietary supplements, URLs seem to be employed as a currency to legitimate one's claim and establish authority (Jacobson et al., 2016; Savolainen, 2014; Wikgren, 2003). In these situations, users are also more likely to question the credibility of the URLs posted by others, especially when persuasive materials are linked to (Savolainen, 2014; Wikgren, 2003).

Importantly, as shown by Jacobson et al. (2016), users with a particular ideology may stick to the sources that support their stances, while both Wikgren (2001) and Savolainen (2014) show that users may appeal to the URLs as representing factual basis or authority for their claims. However, these studies do not address the questions of how users with opposing stances or ideologies employ evidence from online sources for their stance-taking with each other. Furthermore, as pointed out by Polletta et al. (2009) and Wikgren (2003), as well as Colaric & Jonassen (2001), users, either the posters or audience, may not read hyperlinked online information in detail. Therefore, they may undermine their deliberation if they simply attribute authority to the presence of URLs as evidence without investigating the hyperlinked content. Wikgren (2001) proposes that although scientific sources are fact-based, users might actually share them for affective reasons, rather than because of the content. Polletta et al. (2009) further speculate that the glut of information online can undercut personal accounts, or the overwhelming online sources may mean they are perceived as less authentic than personal accounts.

Overall, these studies show that URLs are one of the sources that users refer to in online discussions. URL-posting is mainly for information sharing, and to a lesser extent for stance-taking

when users discuss controversial issues. At the same time, these studies attest to the possibility of examining users' discourse practices in online discussions to understand URL-posting in online spaces, and may provide insights into URL-posting on other online spaces such as social media.

Building on these studies, this thesis examines URL-posting as information sharing practices but pays more attention on its evidencing function in stance-taking. This thesis also fills a gap in the MOOC and education research because thus far there is only one study examining URL-posting in online learning settings (Gallagher & Savage, 2016). Even in that study, only the number of URLs posted was examined without consideration of the discourse practices of URL-posting. This is despite the caution put forward by Colaric & Jonassen (2001) that learners might not assess the relevance of web resources provided by their educators or peers in online learning.

3.7 Theoretical concepts for exploring dialogic nature of online discourse

Two related theoretical concepts – dialogic space and intersubjectivity – are adopted to direct the analysis of users' discourse in online discussions in this thesis. Dialogic space and intersubjectivity also define what I call a *dialogic conversation*, which may or may not arise from user-user interactions depending on users' discourse. These two concepts clearly illustrate the dialogic nature of language underlying social interactions, and thus will further shed light on discourse practices that are facilitative of inviting replies and sustaining conversations with others in online discussions. Furthermore, the two theoretical concepts have also been applied in educational settings, thereby allowing this thesis to achieve its aim of examining online discourse, while acknowledging the educational nature of the research setting – MOOC online discussions.

3.7.1 Dialogic space

Dialogic space has been proposed by researchers from both linguistics (Martin & White, 2005) and education studies (Mercer, 2004; Wegerif, 2010). They share a similar conceptualisation but with a

different emphasis. Both have posited that dialogic space is where uncertainty and multiple voices are welcome, and developed their theory based on Bakhtin (1981). However, Martin & White (2005) conceptualize it as a space to engage and entertain others' voices, whereas Mercer (2004) and Wegerif (2010) a space for inter-thinking among learners. As mentioned earlier, this thesis takes the former assumption such that the language data is considered as evidence of users doing things online, rather than as inference for thinking. Before detailing the works of both groups of researchers, the following first introduces Bakhtin's (1981) work on the *dialogic* nature of human language and *heteroglossia* on which the concept of dialogic space is based.

3.7.1.1 Dialogic language use and heteroglossia

According to Bakhtin (1981), an utterance, whether written or spoken, is normally constructed as if it is addressed to someone, real or imaginary such that it takes up what has been said or established before, and/or generates or anticipates responses. This is the dialogic nature of human language. The relation created between different utterances or texts is *intertextuality* (Fairclough, 2003). At the same time, an utterance that is *heteroglossic* normally entails alternative voices, although the speaker or writer sometimes only explicitly mentions their own voice. For example, in "Probably this policy does not work for us", "probably" implies there is still a small chance that the policy works, which is another voice, while the negation suggests that the utterance is dialogic in response to a prior utterance or context regarding "this policy". An utterance could be framed differently by using different lexical choices and grammatical constructions to enact different degrees of dialogic nature and heteroglossia.

Both the dialogic nature of language and heteroglossia can also be translated from utterance level into the wider context of online discussions. In online discussions, different ideas, i.e., heteroglossia, can be explored, refined and challenged by users such that a dialogic relationship between different ideas can be established. This differs from monologic expressions of views by users without considering others' views (Dahlberg, 2001; Freelon, 2015; Friess & Eilders, 2015). Both

the dialogic and heteroglossic nature of language are the basis of dialogic space proposed in both linguistics and education, as explained below.

3.7.1.2 Linguistic perspective

According to Martin and White (2005), a dialogic space is the communicative context created by an utterance and can be expanded or contracted, depending on whether voices other than the writer's or speaker's are welcomed. These voices can be those that have been mentioned before or prevail in the socio-cultural context, or those that have not yet been raised possibly by potential audiences. A dialogic space is contracted when an assertion is made without the consideration of other voices, or disclaiming opposing opinions. These assertions are normally categorical, sweeping generalizations, 'taken-for-granted' beliefs or matter-of-fact statements that exclude or ignore any other possibilities or voices from others. In contrast, a dialogic space is expanded when a writer or speaker states their own subjective viewpoint as one of the possibilities rather than asserting it, while entertaining alternative voices from others, thus achieving heteroglossia.

The expansion of dialogic space can be achieved by discourse practices utilizing linguistic features such as modals or hedges (e.g., *might*, *probably*, *I guess*) to qualify or mitigate one's own propositions (Biber et al., 1999). By mitigating one's subjective viewpoint as uncertain or as one of many possibilities, the face-threat imposed on other interlocutors who have different views is reduced such that they will be more willing to voice their opinions and engage in the interaction (Brown & Levinson, 1987). Dialogic space can also be expanded when the writer or speaker attributes their argument to outside sources, evidence or hearsay rather than merely mentioning their own idea or framing it as authoritative. One such attribution to outside sources is URL-posting, which is investigated in Chapter 9.

The role of language use in dialogic space has been well exemplified in two studies. In an experimental study on online discussions, Concannon, Healey, and Purver (2017) reveal that a proposition framed in uncertainty (e.g., "Do you think ..." vs. "I think ..."), which allows for

alternative voices, increases the likelihood of deliberation of multiple opinions among the responses in a thread. This could also be the case with expository questions which are open-ended and thus invite others' responses (Martin & White, 2005). In another study on maths classroom discourse (Wagner & Herbel-Eisenmann, 2008), *just* is found used frequently by both teachers and students for various meaning and communicative functions. However, when teachers use *just* to ask students to solve maths problem, for example, "*Just multiply straight across*", this frame the problem as a simple thing to do to students, who in an interview express their loss of agency to suggest and employ other ways of doing the maths. These two studies clearly illustrate the role of linguistic features and discourse practices in expanding or contracting a dialogic space.

The concept of dialogic space has also been applied in online discussions in TED commenting space. Drasovean & Tagg (2015) show that users entertain others' voices with hedges and appreciation before disclaiming and voicing their own view in their comments. The authors suggest that these comments are heteroglossic, as opposed to monologic, and framed in a positive tone such that interactions in this space are not antagonistic despite people having different views. Although this study successfully reveals the dialogic nature of online discourse, it does not explore the difference between posts that receive replies and those that do not, or the turn-by-turn interactions within a thread. This thesis will expand on this aspect while applying the same concept in online discussions.

It is worth noting that the potential effects of linguistic features in expanding or contracting dialogic space vary with the situated social context (Baumgarten & House, 2010; Kärkkäinen, 2003; Martin & White, 2005; Pöldvere, Fuoli, & Paradis, 2016). There is no one-to-one mapping between linguistic features for *dialogic expansion* or *contraction*. For example, in contrast to Concannon et al. (2017), Sotillo & Wang-Gempp (2016) find that *Do you think* are used in an online forum to challenge others rather than invite responses, while Pöldvere et al. (2016) find that *I think* can sometimes expand rather than contract dialogic space. The fact that the function of linguistic features differs according to context highlights the importance of understanding language-in-use in context, that is

the discourse practices. Therefore, in this thesis, the linguistic features found in the analysis will be interpreted based on the actual language data to understand the discourse practices for expanding dialogic space.

3.7.1.3 Educational perspective

Dialogic space as proposed in the educational field can be traced to the dialogic learning approach in classroom proposed by Mercer (2004) who sees the use of language as a means for *inter-thinking*, i.e., *social mode of thinking within peer interactions*. He proposes three types of talk to characterize the unfolding interactions among learners: (1) *disputational talk* where learners disagree with each other with assertions rather than pooling ideas together or trying to achieve mutual understanding; (2) *cumulative talk* where learners agree with each other uncritically and typically repeat what has been said, although sometimes they contribute further elaboration; and (3) *exploratory talk* where learners engage with each other's ideas by evaluations, justifications or counter-arguments, and seek to explore various opinions, without being competitive.

Typically, in *disputational talk* and *cumulative talk*, learners only articulate short expressions such as "No, I don't agree", "Yes, I agree", and the talk can be short-lived. These two types of talk are similar to *dialogic contraction* in Martin and White's (2005) terminology. In *exploratory talk*, alternative ideas are explored with explicit accountability and reasoning, which is similar to what is needed in online deliberation (Freelon, 2015). The exploration of different ideas is similar to Martin and White's *dialogic expansion*, although Mercer (2004) emphasises reasoning in exploratory talk.

Researchers have explored dialogic learning in adult online learning. For example, Littleton & Whitelock (2005) find that postgraduate students often engage in cumulative talk to share information and ideas in online discussions. In contrast, exploratory talk is infrequent and typically takes place in extended threads where learners challenge and counter-challenge. They also find that in exploratory talk, learners do not necessarily employ reasoning but express uncertainty instead, for example, "Just some ideas which may or may not be of help". Response to a question can be

another question that gives a hint to the solution instead of a definitive answer. Their findings suggest that the three types of talk, although first developed for understanding classroom discourse, is also suitable for understanding online discussions, especially those with an educational nature, and the discussion threads can be conceptualized as episodes of talk. Their study also highlights that exploratory talk is not limited to reasoning and more research could be conducted to explore other discourse practices that learners can employ to engage in exploratory talk. This is especially important given that, exploratory talk is relatively rare compared to cumulative talk, while learners seem to avoid disputational talk by not raising alternative views in the online discussion. This study, as well as other studies, suggests that online users, even those in the higher educations, may need to be made aware of how to engage with each other for exploratory talk, especially when they have different stances (Laflen & Fiorenza, 2012; Littleton & Whitelock, 2005; Paulus, 2006).

Wegerif (2010) further extends Mercer's (2004) dialogic learning by proposing that dialogic space is created by exploratory talk. Dialogic space is an interface for inter-thinking and can be full of uncertainty, multiplicity and open-endedness such that learners can admit their uncertainty, ask for advice and be free to change their mind after negotiation with others. Similar to Littleton and Whitelock (2005), Wegerif (2010) also argues that exploratory talk is not just about reasoning, as Mercer (2004) originally proposes, but different voices such that the dialogic space can be *broadened*. Wegerif (2010) emphasises the use of technology and learning design to shift the cumulative talk and disputational talk, that do not explore alternative voices, towards exploratory talk. For example, prompts can be used to interrupt the on-going discussions such that different voices can be considered to broaden the dialogic space, whereas the design for "fast and furious" (Wegerif, 2010, p.313) engagement, which is aimed at completing the tasks, closes down the dialogic space. It is possible that prompt-focused posting in online discussions (Herring, 2013) and teachers' use of *just* to instruct students as reviewed earlier (Wagner & Herbel-Eisenmann, 2008) are similar to this "fast and furious" engagement.

3.7.1.4 Integrating two perspectives to apply dialogic space in online discussions

The linguistic perspective and educational perspective share one key feature in their conceptualization of dialogic space; that is, a space where multiple voices are welcomed and in dialogue, rather than each voice standing by itself in the space. It is worth noting that voicing differences is not enough, as in the case of disputational talk, but they must be in a dialogue, as in exploratory talk. Therefore, it is important to entertain other voices in one's utterances to expand the dialogic space, rather than disregard other voices to contract the dialogic space.

In online discussions, the textual evidence for dialogic space is best observed at the level of threads. Firstly, as Mercer (2005) and Littleton and Whitelock (2005) have shown, the different types of talk are best viewed in episodes of talk. Secondly, given that an utterance, i.e., translated to a comment in online discussions, can expand or contract a dialogic space (Martin & White, 2005), it can be assumed that when a post or reply expands the space, a thread will continue developing, whereas when a post or reply contracts the space, a thread will not be initiated and will be terminated. Therefore, dialogic space will be used as a theoretical concept to direct the analysis of the linguistic features and discourse practices in initiating posts, independent posts, and replies, thus extending previous studies (e.g., Arguello et al., 2006; Crook et al., 2016) that only examine linguistic features without an interpretation on how they are employed for discourse practices that can potentially invite replies and start a conversation.

3.7.2 Intersubjectivity

3.7.2.1 Definition as integrated from education and linguistics

Literally, intersubjectivity refers to the *interrelationship* between the *subjectivities* of different individuals (Kärkkäinen, 2006). The concept was developed in philosophy, cognitive sciences, linguistics and sociology for understanding human cognition and social interactions (Du Bois, 2007). In this thesis, I draw on the conceptualization of intersubjectivity in education and linguistics to further explicate the relationship between different voices within a dialogic space; that is, how users engage with each other's voices within a thread and how their replies build on each other's.

In education research, intersubjectivity is the desirable outcome of shared understanding and joint meaning-making achieved by learners through peer interactions (Dennen & Wieland, 2007; Hall, 2010; Mercer, 2004; Stahl, 2015; Wegerif, 2010). In linguistics, intersubjectivity is a process by which interlocutors engage with others' voices and position themselves in relation to others, as demonstrable via their language (Du Bois, 2007; Martin & White, 2005). The former can be said to be more focused on the outcome of interactions, whereas the latter draws attention to the processes.

Both perspectives share a similarity in suggesting that intersubjectivity is not just interactions. In other words, intersubjectivity contrasts with parallel monologues, where interlocutors seem to be in a dialogue or to be replying within a thread, but do not entertain others' voices such that there is no shared understanding and acknowledgment of each other's views. Similarly, group or peer discussions are not necessarily a dialogic space if different voices are only raised but not in a dialogic relationship, and each voice is not explored nor acknowledged (Mercer, 2004, Wegerif, 2010). This differentiation of dialogic conversations from interactions is also deemed important in both online deliberation and online learning (Dahlberg, 2001; Hall, 2010). Integrating these two perspectives, intersubjectivity can be defined as the co-constructive process of building on each other's contributions in a conversation to achieve shared understanding.

To achieve this shared understanding, this thesis holds that differences, misunderstandings and breakdowns can be negotiated and repaired by interlocutors in their interactions, while similarities and common ground can be acknowledged (Marra, 2012; Schegloff, 1992). Thus, this thesis will focus on users' discourse practices that could help them to engage in the processes of intersubjectivity in online discussions by exploring how they respond to each other within a thread. To this end, intersubjectivity in stance-taking as conceptualized by Du Bois (2007) is used to direct the analysis of threads in this thesis because stance-taking is dialogic in nature and happen when users engage in information exchange, socialization or deliberation in online discussions.

3.7.2.2 Stance-taking

According to Du Bois (2007), stance is a person's subjective opinion towards a particular issue, object or proposition (e.g., "I think"). The stance expression can be affective (e.g., "I am glad it turned out well!"), evaluative (e.g., "That is horrible!") and epistemic (e.g., "I know.") in nature, and can be implicitly invoked or explicitly conveyed. A stance can then be aligned or disaligned with by others in a conversation. The (dis)alignment roughly corresponds to agreement or disagreement between stance-takers. However, the alignment or disalignment is not categorical but a continuous (re-)calibration of the interrelationship between the different stances. Intersubjectivity involves the constant updating and negotiation of each other's positions along this continuous scale of stance (dis)alignment on a turn-by-turn basis in the interaction. In other words, by taking a stance towards the same objects, the interlocutors are positioning themselves in line with or against each other. In this way, intersubjectivity is dialogically and sequentially co-constructed (Du Bois & Kärkkäinen, 2012; Schegloff, 1992).

During a disagreement, users can be positioned at extreme ends of the scale of stance at the start, but through their discussions with each other, their positioning could be updated and moved towards either agreement or a position that reaches mutual understanding of differences (Marra, 2012; Nathan, Eilam, & Kim, 2007). In a maths classroom, Nathan et al. (2007) found that breakdowns in students' group discussion for problem solving can be due to their different interpretative frames and representations. For example some students view a pie literally, whereas others view it geometrically. Through their discussions, with gesturing and drawing, students can understand where each other comes from, i.e., their different representations, thus reaching mutual understanding while taking into account each other's ideas.

Stance-taking can also be translated into the three types of talk proposed by Mercer (2004), where cumulative talk equates to alignment, disputational talk disalignment, and exploratory talk takes place where intersubjectivity is negotiated. Besides stance-taking, the constant negotiation of each other's stances may also involve reflection on assumptions, clarification and repair of

misunderstanding as a conversation evolves (Nathan et al., 2007; Schegloff, 1992; Wegerif, 2010). Therefore, Du Bois & Kärkkäinen (2012) also conclude, intersubjectivity is best observed in episodes of talk, as in dialogic learning (Mercer, 2004).

Although it is hard to pinpoint specific linguistic features involved in intersubjective processes, given that they have to be contextualized to the conversations, researchers have identified several discourse practices that are observed to be facilitative of intersubjectivity. These include probing with questions for clarification, acknowledging contributions by others and situating others' contributions within one's own contributions, making rationales explicit (Dennen & Wieland, 2007), and using parallel and resonating structures to indicate involvement (Kärkkäinen, 2007), as well as non-verbal practices such as gesturing and drawing (Nathan et al., 2007). Based on the previous literature, three discourse practices – epistemic stance, meta-communication and identity performance – are identified as particularly relevant for intersubjectivity in online discussions because of their role in oral conversations and online discussions (Grabill & Pigg, 2012; Heritage, 2012; Tanskanen, 2007), and are explained in the following subsections.

3.7.2.3 Epistemic status and stance

According to Heritage (2012), epistemic status refers to one's knowledge and understanding, i.e., the status of "knowing", "partially knowing" and "unknowing". Interlocutors can position themselves as being either higher or lower in status in relation to their interlocutors in a conversation. Positioning oneself as "knowing" is to claim one's expertise, authority or legitimacy (Bellander & Landqvist, 2020). Epistemic stance is the expression or marking of epistemic status, and it can assist users' calibration of the interrelationship between different stances, thus facilitating processes of intersubjectivity.

As concluded by Biber et al. (1999), epistemic markers are used to indicate one's certainty or doubt (e.g., "definitely", "possibly", "typically"), limitation, sources or perspective of knowledge (e.g., "according to", "scientifically speaking", "there was a suggestion..."), and commitment to a

stance (e.g., “I know”, “I am sure”, “it is essential”). Besides epistemic markers, language users also express their stances with declarative (e.g., “You are married.”) or interrogative morphosyntax (e.g., “Are you married?”, “You’re married, aren’t you?”) (Heritage, 2012, p. 6). By indicating epistemic stance, interlocutors can co-construct each other’s epistemic status, thus achieving shared understanding and intersubjectivity. All these markers and ways of expressing epistemic stance need to be situated within the context, i.e., the preceding contributions and the forthcoming contributions, to understand how interlocutors build on each other’s turns for stance-taking.

3.7.2.4 Meta-language

Meta-language is an umbrella term used in this thesis to refer to users commenting on their own or others’ act of commenting, rather than mentioning the content or topic of the discussion. Other similar terms and concepts include meta-comments (Myers, 2007), metapragmatic expressions (Kleinke & Bos, 2015; Liu & Liu, 2017; Tanskanen, 2007), meta-discourse (Benwell & Stokoe, 2002; Stahl, 2015; Sutherland, 2015), and meta-talk (Swales, 2001). Examples of meta-language include referring to one’s or the other’s posting or connections between them (e.g., “To reply to Harriet, I would say that I think that my comments were somewhat misunderstood. ...” (p.91), commenting acts (e.g., “Just wanted to clarify”, p.92; “Sorry for the rant”, p.100), language use (e.g., “I know this is inflammatory”, p.91) (Tanskanen, 2007), and emphasising (e.g., “My point is”, “The thing is”, Swales, 2001).

Meta-communication has been found to be important in dialogic learning (Sutherland, 2015), group collaborations (Stahl, 2015), and creating common ground for intersubjectivity when interlocutors point out the similarities or differences in their views (Liu & Liu, 2017). More importantly, in asynchronous online discussions, the persistence of the text allows users to reflect on what has been put down by themselves and others in the discussions, thus facilitating meta-linguistic comments (Herring, 1999; Lapadat, 2007; Tanskanen, 2007; Wegerif, 2010). Meta-language may reveal what users see as appropriate or important in a discussion and help compensate for the

lack of non-verbal social cues with which to manage their interaction in online discussions, given that meta-language involves overt comments on commenting acts (Kleinke & Bos, 2015; Myers, 2007; Tanskanen, 2007). This explicitness of meta-language facilitates shared understanding, thus intersubjectivity.

The analysis of metapragmatic expressions in mailing lists and message-board discussions conducted by Tanskanen (2007) is reviewed in depth next because it is relevant to how users manage their interaction in online discussions. Tanskanen (2007) finds that the most frequent metapragmatic expressions concern the appropriateness of users' own postings, either other-initiated, i.e., in response to others' replies to their previous comments (e.g., "As the sender of the original post, I do want to clarify that I did not intend to suggest that", p.92), or self-initiated, i.e., pre-empting possible responses from others (e.g., "At the risk of adding further to the list's exceptionally heavy mail volume, I think I should try to explain why I think this discussion", p.95). These metapragmatic comments are an indication that users repair or pre-empt miscommunications or misinterpretations, suggesting they consider others' perspectives and attempt to engage in intersubjectivity. As the discussion of this study shows, metapragmatic expressions are dialogic in nature (Bakhtin, 1981) because they address actual audiences in the case of other-initiated metapragmatic posting, and imagined or potential audiences in the case of self-initiated metapragmatic posting.

Besides repairing and pre-empting, metapragmatic expressions are used to control the flow of the discussions (e.g., "Well, you've certainly ended this discussion effectively. All that's left to say is 'I rest my case.'", p.100) or feedback on the threads (e.g., "we seem to be spiralling down into a general discussion of", p.101) (Tanskanen, 2007). In these cases, users explicitly mention their judgments on individuals' posting behaviours or on-going discussion in the threads, revealing what users may consider as appropriate in the discussions. The metapragmatic discussions also suggest that communicative norms in an online discussion are fluid and co-constructed by users, rather than

being solely dictated by the aim of the discussions set up by the hosts or designers (Kleinke & Bos, 2015; Tanskanen, 2007).

3.7.2.5 Identity

Most online discussions are anonymous. Users typically do not know each other in person and lack knowledge of each other's background. The only explicit information about each user is typically the profile picture and the often limited self-description in the profile. Given this lack of personal background information, the main way users can perform their identity or get to know others is via their postings which, in a text-only context, happens through discourse. This identity performance differs from the fixed characteristics or demographic background of a person, but is a positioning of self in relation to others. This observation draws on a wider conceptualisation of identity as discursively constructed, emerging from interaction with others and change with context (Benwell & Stokoe, 2006; Bucholtz & Hall, 2005; Jones & Hafner, 2012). The information revealed through users' identity performances may construe their perspectives in stance-taking, thus facilitating shared understanding and intersubjectivity among users.

Identity performance in online discussions has been found to be a way to legitimate one's contribution in information exchange and stance-taking. For example, in TripAdvisor hotel reviews, Vásquez (2018) found that, although personal identity is not related to the hotel being reviewed, users typically preface their assessment of a hotel with their personal information. The information includes financial ability and reason for travelling so as to draw the attention of readers with a similar background. Users also preface their negative assessment with an emphasis on their experience in travelling ("In my 25 years of business travel", p.73), or even an explicit mention of them being not unreasonable ("Usually not one to complain", p.77), which discursively constructs themselves as a reasonable and experienced traveller. These identity performances establish their authority in understanding hotels and also contextualize why they think a particular hotel is worse than their normal expectation based on wide experience. In other words, identity performance

reveals one's subjectivities to others. Vásquez (2018) concluded that this identity performance not only helps other readers to decide which reviews to heed but also provides justification to the reviews as credible.

Identity information can also be used to execute one's agency in directing the framing of the on-going discussions. In a thread in a blog roll hosted by a museum, Grabill and Pigg (2012) found that users enact different identities to frame the on-going discussions on cervical cancer, from judgement towards young women to information exchange. The first few users who talk about young women include a woman "who knows this information" (p.110) while differentiating herself from those "young women" (p.110) who do not know, and a mother whose "two little girls ...might die...from a completely preventable disease" (p.110). However, the talk about young women switches course when a young woman performs her identity ("Yet I have to say that as a seventeen year old girl ...", p.110) while explaining her point of view on the issue - that the information on cervical cancer is not readily available for young women. As evidenced by the posts afterwards, her post not only invites other seniors to provide information, but also creates a space for other similarly aged people to join the discussion. After this post, more young girls identify themselves by signing off with their ages and voicing their opinions, especially those infected with HPV.

This discourse strategy of identity performance is particularly significant in this context because the young women are first construed by others as passive and ignorant in preventing cervical cancer, but by performing their identity, they execute their agency as young women who take stance and engage in information exchange (Grabill & Pigg, 2012). This identity performance also contextualizes the questions posted by them and moves the discussion from abstract to concrete scenarios. Although the authors do not conceptualise this identity performance as intersubjectivity but as a way of framing on-going discussions, their analysis shows that identity performance makes explicit what one's argument is based on, thus facilitating the interactions between the young women and those talk about them. This suggests that identity performance also facilitates shared understanding and intersubjectivity in this and other online discussions.

3.8 Importance of sustained conversation for exploratory talk and intersubjectivity

As revealed in Table 3.1 earlier, initiating posts that receive replies do not necessarily develop into sustained conversations. Although the differentiation between short-lived and sustained conversations can be arbitrary in terms of the exact length of threads, it can be argued that one-reply or two-reply threads commonly found in online discussions (Beth et al., 2015; Napoles et al., 2017; Tubman et al., 2016; Cui et al., 2017) are not sufficient for an exploratory talk or processes of intersubjectivity. Assuming that exploratory talk consists of reciprocal critique and exploration of multiple voices, short-lived threads may be too short for all these to happen. Similarly, to achieve intersubjectivity, users might need to engage in turn-taking to clarify and update understandings of each other's voice, which necessarily entails more than one reply from at least two interlocutors to allow ideas to unfold and build up. This speculation can be confirmed by previous research that illustrates exploratory talk and intersubjectivity in episodes of face-to-face conversations where interlocutors engage in sustained conversations (Kärkkäinen, 2003; Mercer, 2004; Schegloff, 1992) and long threads in online discussions (Grabill & Pigg, 2012; Jaworska, 2018).

For example, in an online discussion of a hiking backpacker group investigated by Ziegler et al (2014), a thread evolves to 66 replies within nine days, with several users replying a few times. This thread is initiated with a post containing a question "How do you carry your ground coffee?" (p.69) in which the user shares their experience of preparing for a hike. The first three replies contain suggestions asking the user to use a bag and construing the issue as trivial, e.g. "just use" and "should be fine" (p.71). This might look like an information exchange that has been resolved. But, through micro-analysis of the replies within the thread, the researchers find that users also explore new ideas and engage in negotiation besides providing answers. One of the replying users questions the user "What will you do with the used coffee grounds?"(p.71). This question sets their position differently from the trivial issue of "coffee" and starts a full-blown negotiation on "leav[ing] no trace" (p.73) during hiking. In the discussion ensued, users question each other's assumptions or

their positions during hiking while making explicit their own, e.g., hiking ethics vs. one's convenience, and explore different issues arising from carrying coffee, including the different effects of throwing or burning coffee grounds. Exploration of different positions and voices would not be possible if the thread stops at the first three replies. The expansion of dialogic space is also evidenced by the voicing of new issues, rather than sticking to the initial positioning of preparing for a hike. This thread also shows that both information exchange and negotiation for intersubjectivity can happen at the same time.

From this example and other micro-analyses of threads in online discussions (Jaworska, 2018; Littleton & Whitelock, 2005; Ziegler et al., 2014), it can be argued that exploratory talk and intersubjectivity are more likely to occur in long and sustained threads, rather than short-lived threads. However, it is important not to undervalue short threads in online discussions. Short threads can include cumulative talk through which users express agreement or gratitude to other users who have voiced similar views (Baym, 1996), or question and answer in information sharing (Poquet et al., 2018). Therefore, short threads can still be considered potentially valuable for socialization and information seeking, compared to not receiving replies.

3.9 Importance of disagreement for exploratory talk and intersubjectivity

Disagreement has been seen as a double-edged sword in the literature on online discussions and deliberation, as well as face-to-face interactions. Compared to exploratory talk, disputational talk can shut down a dialogic space and hinder dialogic learning (Martin & White, 2005; Mercer, 2004; Wegerif, 2010). However, disagreement also means that different opinions are voiced, instead of an echo chamber where there is only one voice (Freelon, 2015; Walter et al., 2018). Online deliberation can be achieved if users engage with each other's voices (Dahlberg, 2001; T. Graham et al., 2016; Landqvist, 2016; Lewiński, 2013). Furthermore, disagreement, as well as agreement, is an indication of interactivity and responsiveness, because disagreement is always in response to a prior message

or content presented in the immediate communicative context, either by other speakers or users, or the host of a website (Bakhtin, 1981; Baym, 1996; Bolander, 2012; Keisanen, 2007; Kleinke, 2010; Lapadat, 2007; Pomerantz, 1984). Therefore, disagreement is one indication of user-content or user-user interactions, and it is important to understand its role in online discussions, and how it might be related to dialogic space and intersubjectivity.

The value of disagreement has been illustrated in both face-to-face and online discussions. In daily oral conversation, incompatible positions between speakers motivate information seeking to facilitate reconciliation, suggesting that disagreement can trigger a negotiation process (Robinson, 2009). Similarly, in online discussions, clarification and co-constructions following a disagreement could drive the discussion forward (Concannon & Healey, 2015), and disagreement with elaboration can help introduce new materials or directions into the discussion (Baym, 1996). Research from classroom learning also shows that disagreement can cultivate learners' ability to achieve shared understanding and express divergent views despite not converging to a single view, and foster sustained conversations (Nathan et al., 2007). All these findings attest to the potential of disagreement for generating processes of intersubjectivity that will result in co-construction and exploratory talk, rather than disputational talk.

However, this potential seems to be conditional on participants' discourse. For example, in classroom learning, arguing in order to reach consensus is more effective for knowledge co-constructions and exploratory talk compared to arguing to defend oneself or defeat others (Felton, Garcia-Mila, Villarroel, & Gilabert, 2015; Mercer, 2004). Polite disagreement, evaluation and justification increases the likelihood of problem solving in the maths classroom (Chiu, 2008). These studies highlight the importance of language use and discourse practices in discussions, especially its role in shifting disputational talk to exploratory talk.

In online discussions in distance learning, some learners, however, tend not to challenge others' ideas, probably to avoid the potential confrontation arising from disagreement (Littleton & Whitelock, 2005; Paulus, 2006). This can be because they are unfamiliar with online discourse for

debating without compromising their relationship with others, especially when they do not know each other in person. Yet, avoiding or ignoring disagreement can be a missed opportunity to be included in a community (Marra, 2012), or to be involved in a sustained conversation that might lead to negotiation, intersubjectivity, exploration or reconciliation (Bou-Franch & Garcés-Conejos Blitvich, 2014). In contrast, Baym (1996) suggests that online disagreement to challenge ideas might be easier to express than in face-to-face conversations given the distance between users. For example, users use fewer hedges and sound less polite in online discussions compared to face-to-face discussions (Brennan & Ohaeri, 1999). This means that disagreement could also easily lead to disputational talk, impoliteness, standoff or incivility, which are quite common in online spaces (Bou-Franch & Garcés-Conejos Blitvich, 2014; Kleinke, 2010; Kleinke & Bos, 2015; Sotillo & Wang-Gempp, 2016).

Therefore, it is important to examine how disagreement can be utilized by users for co-construction, instead of developing into disputational talk or users simply leaving the discussion. This will also raise users' awareness of the necessary language practices to engage in online discussions more effectively (Laflen & Fiorenza, 2012; Littleton & Whitelock, 2005; Paulus, 2006). As argued by Marra (2012), being able to express disagreement and engage in negotiation can be an interaction achievement in a community, although different communities may have different norms of how to disagree. The discourse practices involved in turning a disputational talk or disagreement into exploratory talk can be understood in terms of intersubjectivity where interlocutors adjust to each other's position and update themselves on each other's epistemic status. This will be examined in this thesis with micro-analysis of threads, where users' discourse practices in their replies to each other are investigated.

3.10 Conclusion

Overall, the literature review points to the need to examine discourse practices that can initiate and sustain dialogic conversations in online spaces, as well as how users employ URL in their stance-

taking. Research in online discussions, especially those examining sustained threads, as well as face-to-face discussions, have shown that users' discourse does influence the development of conversations, including in the case of exploratory talk, and turning disagreement into meaningful discussions. However, the extent of lack of replies and short-lived interactions in online discussions suggest that such conversations do not happen often. Although there has been an endeavor in using technology to promote conversations, there is also a need to raise users' awareness of their agency and discourse practices in online spaces, especially when users hope to receive emotional or informational support from others, deliberate issues with others or simply socialize for leisure.

There have been studies revealing linguistic features that can potentially increase the chance of receiving replies in online discussions, attesting to the fact that there are indeed differences in language use between posts that receive replies and those that do not. Therefore, it is important to differentiate between initiating posts and independent posts, unlike previous studies on MOOC and online discussions that do not make such a differentiation. However, the linguistic features need to be situated in context to reveal the discourse practices that enact social relationships, including initiating and sustaining conversations with others.

Therefore, in this thesis, to interpret the linguistic features and discourse practices in online spaces, I employ two theoretical concepts – dialogic space and intersubjectivity. Both concepts focus on the dialogic nature of language, and are useful in distinguishing dialogic conversations from parallel monologues in user-user interactions. This distinction is important for understanding online deliberation as well as information exchange and social interactions. Dialogic conversations are attested by various studies, either on online spaces or spoken conversations, that show that discourse practices can expand or contract dialogic spaces and facilitate processes of intersubjectivity. These various studies also provide evidence for the assumption that language achieves things in the real world.

The literature review also shows that in-depth analysis of discourse practices and micro-analysis of threads are needed to examine users' online discourse to better understand how a

dialogic conversation unfolds. The online spaces that have been examined with analysis directed by dialogic space or intersubjectivity include TED and museum blog commenting space, small-scale online learning discussions, mailing lists, interest group discussions. As far as I am aware, this might be the first time MOOC online discussions is examined from this perspective in a large scale, although Collins' (2019) corpus analysis might have touched upon the interactions among users.

Alongside language, URLs are also one way of expanding dialogic space to voices outside of the immediate online discussions. Although most studies find that users post URLs to different types of sources for information sharing, there are times that users employ URLs as evidence for stance-taking. This latter aspect of URL-posting may underlie their intersubjective process, especially when they disagree over an issue. However, there has been limited in-depth analysis of discourse practices surrounding URL-posting in this aspect.

In short, this thesis aims to extend previous studies by systematically characterizing the linguistic features and derive discourse practices in initiating posts, independent posts and replies to explore how users can employ their language and URLs in online spaces to establish dialogic conversations with others. This will be achieved by a corpus linguistic approach combined with micro-analysis of threads, which I turn to in the next chapter.

Chapter 4

A corpus linguistic approach to online discourse

4.1 Introduction

This chapter introduces the main methodology - corpus linguistic approach. This choice of methodology is based on my position that language does things, thus users' textual contributions constitute evidence for their social actions in online spaces. The rationale of approaching online discourse with corpus linguistics is explained, with a literature review of relevant corpus studies. In this thesis, corpus linguistics is combined with micro-analysis of threads that draws on selected principles of conversation analysis to explore the online discourse in the MOOC online discussions. More specifically, the linguistic features and discourse practices of initiating posts, independent posts and replies will be explored with the corpus methods of keywords, analysis of concordance lines and collocates, while the replies within threads will be further zoned into with micro-analysis of threads. Lastly, the practice of URL-posting will be explored with both corpus methods and micro-analysis. Table 4.1 presents an overview of the methods used in this thesis.

Table 4.1 Methods used in this thesis

Linguistic features and discourse practices in:	Corpus linguistics	Micro-analyses	Chapters
Initiating posts	Keyword Analysis, concordancing, collocation analysis	-	6
Independent posts	Keyword Analysis, concordancing, collocation analysis	-	6, 7
Replies	Keyword Analysis, concordancing, collocation analysis	Three selected threads and threads in which <i>agree to disagree</i> takes place	8
URL-posting	Concordancing, collocation analysis of the word <i>link(s)</i> and URL addresses.	Users' posting and co-construction of the value of URLs	9

As will be explained in this chapter, these methods are employed to address the research questions:

RQ1: What are the differences in the linguistic features and discourse practices that regularly occur in

- initiating posts that receive replies and start a discussion thread,
- independent posts that do not receive replies,
- replies, especially those in sustained discussions

RQ2: How do these discourse practices initiate, sustain or hinder dialogic conversations in online discussions?

RQ3: How does URL-posting initiate, sustain or hinder dialogic conversations in online discussions?

4.2 Position towards textual contributions in online spaces

This thesis examines the dialogic nature of online discourse, specifically how users employ language to initiate and engage in conversations with others in online spaces. In other words, this thesis follows the general assumption of discourse analysis which can be termed language-in-action – that language is used as a meaning-making tool by speakers, writers or internet users to enact their identity, construe their social world, and co-construct relationships with others (Fairclough, 2003; Heritage, 2012; Herring, 2004; Vygotsky, 1978). Therefore, in this study, users' textual contributions in online spaces is held to constitute evidence of social practices. This is in contrast to previous MOOC research and other studies investigating posts that receive replies (Arguello et al., 2006; Crook et al., 2016; Kellogg et al., 2014; Wise et al., 2016), that typically assumes texts are static representation of users' inner cognition (Wegerif & Mercer, 1997; Wise & Paulus, 2016).

Researchers in the field of discourse analysis, as well as corpus linguistics and micro-analysis, see language and discourse practices as an active process between interlocutors in co-constructing their social world and relationships, such that interpretations of which vary with the co-text and context (Herring, 2004; McEnery & Hardie, 2012; Wise & Paulus, 2016). For example, the phrase *I think* may be categorized as reflecting cognition in the coding and counting paradigm. However, from the perspective of discourse analysis, it could be used for multiple functions including expressing uncertainty or taking a strong stance, depending on the co-text and context (Baumgarten & House, 2010; Kärkkäinen, 2003; Pöldvere et al., 2016; Simon-Vandenberghe, 2000). Similarly, as shown by Baym (1996), expression of agreement does not necessarily reflect a user's thinking, but can be a mitigation strategy or a strategy to stay coherent in the online discussion. As we shall see in the review of corpus studies on online discourse in section 4.4, various language patterns can be employed for different discourse practices in online spaces. Therefore, I argue that the assumption underlying MOOC research on users' comments – that users' textual contributions are static representation of users' thinking and can be reduced to code – is not sufficient for exploring online discourse in the user-user interactions, and corpus linguistics has much to offer.

4.3 A corpus linguistic approach to online discourse

To analyse users' discourse practices based on the textual evidence in their comments while harnessing the big data available from online discussions, a mixed methodology, corpus linguistics, is adopted in this thesis. A corpus linguistic approach consists of a set of established procedures and methods, including keywords analysis, concordancing and collocation analysis, that can be used in combination to investigate language use in large bodies of textual data. *Keywords* refers to words used significantly more often in the corpus; *concordance lines* show a word of interest in context within lines of texts; *collocates* refer to words co-occurring significantly often with the word of interest or keyword (see section 4.5 for details). These procedures reveal repeated patterns of language usage in the corpus, thus facilitating the observation of discourse practices that are common in the specific language community.

Central to corpus linguistics is the assumption that, with a corpus of a suitable size, recurring (and rare but important) language patterns can be identified and analysed to reveal language usage and language users' construal of the social world and their interactions. This kind of information might not be easily available through self-report or human intuition but require a systematic analysis of a corpus of naturally occurring language data, as opposed to elicited data such as interview (McEnery & Hardie, 2012). For example, by analysing a corpus of spoken business data, Cheng (2004) reveals how hotel staff unknowingly employ discourse practices that lack politeness strategies when interacting with customers, and suggests language awareness training to be incorporated into staff training. This study illustrates that, unlike corpus data, a person's post-hoc account may not reflect how their discourse is actually intended and in particular received by others. Instead, corpus analysis can be employed to identify repeated (or rare yet important) language patterns through which social world is construed.

Although there is no consensus on what defines corpus linguistics, it is generally agreed that a corpus analysis typically involves both quantitative analysis and qualitative analysis (Biber et al., 1999; McEnery & Hardie, 2012). Quantitative analysis is conducted on the frequency data, i.e., the

number of occurrences of words or linguistic patterns in the corpus, typically achieved by keyword analysis, frequency count, and collocation analysis. The qualitative analysis is conducted on the co-text or context where a word occurs in the corpus, typically achieved by concordancing (McEnery & Hardie, 2012). Both analyses are usually conducted in synergy to examine the form and function of language use (Biber et al., 1999). Forms, that is words, or collocates, are the basis of quantitative analysis whereas function is examined by the qualitative analysis. Usually, some kind of discourse analysis is conducted for the qualitative analysis alongside concordancing in corpus studies (Baker et al., 2008). In this thesis, micro-analysis of threads is employed to explore how a dialogic conversation unfolds and how discourse practices influence the development of conversations turn-by-turn (see section 4.7).

The simultaneous quantitative and qualitative analysis of textual data, or form and function analysis, in corpus linguistics is crucial for the present analysis of linguistic features and discourse practices in each type of comments. As reviewed in Chapter 3, most previous studies (Arguello et al., 2006; Crook et al., 2016) examine only the linguistic features through quantitative analysis without explaining how the features are used in context or employed for a particular discourse practice, whereas other studies within the coding and counting paradigm (Chen et al., 2020; Wise et al., 2016) do not reveal users' actual language use. Unlike these approaches, the corpus linguistic approach reveals actual language use and discourse practices, on top of the quantitative analysis of language patterns.

Another central tenet of corpus linguistics is that the corpus not only collects naturally occurring language, but also maintains the actual language data, its co-text and context as fully as possible in the data extraction and analysis (to be illustrate in the case of the current corpus in Chapter 5), especially for concordancing and qualitative analysis. This contrasts with current machine learning techniques used in MOOC research (e.g., Cui et al., 2017) which breaks down text into bags of words and discards function words, such as modals, grammatical words and pronouns which have been well-established as linguistic resources for epistemic expression, stance-taking and

establishing interpersonal relationship (Biber et al., 1999; Carter & McCarthy, 2006; Martin & White, 2005). The different choices in processing and analysing textual data is due in part to different research aims, for example the content of a comment might be more readily reflected by nouns, lexical verbs and adjectives (Wise & Paulus, 2016). The word-based automatic analyser Linguistic Inquiry and Word Count (LIWC, Pennebaker et al., 2001) used by Arguello et al. (2006) and Crook et al. (2016) to examine the linguistic features in initiating posts also employs similar approach. The word-based automatic analyser categorizes words into dimensions that cover both linguistic features, such as pronouns, informal speech and negations, and psychological processes, such as cognitive mechanism and core drives and needs. As argued earlier, qualitative analysis is needed to further understand *how* these bags of words or linguistic features are used in discourse practices for establishing dialogic conversations with others, therefore a corpus linguistic approach is preferred over these approaches in this thesis.

In corpus linguistics, the quantitative analysis is based on the formal structures of language, such as words, collocates, lexical bundles and clausal structures which are observable and indisputable, whereas qualitative analysis involves the interpretation of the functions of these observable formal structures in the text they occur in. It can be argued that the quantitative analysis of formal linguistic structures is rather data-driven and relatively independent of human interpretation, thus less biased compared to the pre-defined codes or semantic categories which are produced by humans or machines in the coding and counting paradigm (e.g., Kellogg et al., 2014; Cui et al., 2017). To some extent, the data for the statistical analysis in the coding and counting paradigm is codes created by qualitative analysis in relation to a priori frameworks (Herring, 2004; Wegerif & Mercer, 1997; Wise & Paulus, 2016). In contrast, in corpus linguistics, the data for statistical analysis is word forms, which is not affected by human interpretation.

In short, this thesis takes a corpus linguistic approach to examine users' discourse practices in online discussions, while extending previous studies that only focus on linguistic features or reduce users' textual contributions to codes, thus achieving the methodological objective set up in

Chapter 1. With the corpus methods, I am able to exploit the big data available from online spaces and identify recurrent language patterns through quantitative analysis, while investigate the discourse practices employed by internet users through qualitative analysis.

4.4 Corpus studies on online discourse

As reviewed earlier, only one corpus linguistic study has been conducted on MOOCs (Collins, 2019). Besides this study, corpus linguistics has also been used to examine discourse practices in various text-based asynchronous online spaces, some of which are reviewed in Table 4.2. These studies move beyond exploring the difference between online discourse and spoken or written discourse (D. Knight et al., 2014; Yates, 1996) by focusing on the discourse practices in a particular online space, for example *medicalizing discourse*, as Hunt & Harvey (2015) term it to describe teenagers' advice seeking in health forums.

Table 4.2 Selected corpus studies on online discourse

Study	Online space	Corpus size	Main aims of using corpus analysis	Main corpus findings
Beers-Fägersten, 2008	Message-board postings on five hip-hop websites	102,343 words	Examine discourse practices employed by users to identify themselves as members of the hip-hop community.	<ul style="list-style-type: none"> -The keyword <i>yo</i> is primarily used as an opening. Another keyword <i>peace</i> has a specific meaning in this community as a way to express salutation or farewell in the closing. -One common pattern of the keyword <i>you</i> is <i>if you</i>, which is employed by users to seek help. However, when <i>if u</i> is used, it is used to challenge others. -Taboo terms are used to construe hip-hop as counter-culture, whereas slang terms are used to enact one's in-group identity. -These language practices suggest that contributors construct their identity by how they write, rather than simply declaring oneself "as a hip hop person"
Hewings, Coffin, & North, 2009	Two discussion forums in distance learning	49,048 words	Use corpus analysis to support and extend findings on discourse analysis of students' interaction.	<ul style="list-style-type: none"> -The keywords <i>I</i> and <i>you</i> are used often with <i>think</i> and <i>agree</i> by students to present modalized claims or support others' claims. This contrast with tutors who use <i>you</i> predominantly for directives. -This suggests students can establish solidarity and community building while discussing topical issues in the same comments, besides those comments that are not on topic.
Drasovean & Tagg, 2015	Commenting spaces in twenty-two TED videos	340,938 words	Use corpus analysis to support in-depth analysis of commenting spaces of three TED videos within the corpus.	<ul style="list-style-type: none"> -The keyword analysis shows that users co-construct affiliation with the same keywords when talking about the topic related to the video. -The keywords found also supports the author's in-depth analysis that shows that users employ discourse practices that show appreciation and relationship-building while deliberating on issues.
Hunt & Harvey, 2015	- Advice-requesting messages sent to a professional-	- 2 million words - 156,000 words	Examine users' concern and personal experience of eating disorder, and how they negotiate anorexic identity.	<ul style="list-style-type: none"> - Analysis of the collocates of the keyword <i>anorexic</i> in the advice-requesting messages reveals what the authors term <i>medicalizing discourse</i>. - The users express uncertainty regarding their conditions, but employ "quantification rhetoric" for their symptoms, and refute a third-party's

	run health interactive website - A support forum			<p>judgment in their advice-seeking. These discourse practices construe anorexia and normality as on the same continuum.</p> <ul style="list-style-type: none"> - Anorexia is also construed as a skill, especially as evidenced by the collocate <i>want</i>, suggesting a pro-anorexia discourse. - In the support forum for those recovering from anorexia, the users' discourse rejects an identity wholly defined by anorexia, but compartmentalizes it as only one aspect of their identity. - Interestingly, in this forum, <i>anorexia</i> is used less often than ED or <i>eating disorder</i>, and the qualitative analysis shows that users avoid using anorexia to define themselves, while construing ED as an external agent that affects them. - Users' discourse in health forums reveals their experience and understanding of their health, and provides insights to healthcare professionals for care management of patients.
McDonald & Woodward-Kron, 2016	Online support group for bipolar disorder	8.4 million words	Examine the shifts in lexico-grammatical and discourse-semantic choices among members when they become more veteran in the group.	<ul style="list-style-type: none"> - When providing advice to new members, the discourse of veteran members are modulated imperatives and modalized statements that mark the source of their knowledge, rather than imperatives with general comments that come off as face threatening. - <i>I would + adjunct</i> constructions are employed by veteran members to assume shared identity with the new members. - The veteran members employ vague language, <i>things</i>, to provide social support when they lack information from the other members. - The findings reveal the experience of patients that are not easy to gain from formal healthcare settings, and provide insights to healthcare professionals for their communications with patients.
Sotillo & Wang-Gempp, 2016	Public bulletin boards for residents of five northern New Jersey towns	46,300 words	Examine discourse practices employed by users of different ideologies in political discussions.	<ul style="list-style-type: none"> - Collocates of the words of interest, the candidates for elections, not only reveals the negative sentiments towards the candidate, but also ad hominin argument because the discourse deconstruct the candidates for other identity, including <i>maid</i>, <i>millionaire</i>.

				<ul style="list-style-type: none"> - Frequency lists show that <i>think</i> and <i>know</i> are used the most when users take stance, but it is a discourse practice that exert certainty and one's knowledge about the situation. - The phrase <i>Do you think</i> and similar constructions are used to challenge rather than open up dialogue. - In-depth analysis of pronoun use reveals users enact in-group and out-group identity in their stance-taking to claim power over the others. - This finding reveals the voice of powerless individuals online and call for attention of electorates to pay attention to democracy communities online.
--	--	--	--	--

The overview in Table 4.2 attests to users' agency in employing discourse practices to enact their identity and relationships with others in online spaces, while highlighting how corpus analysis reveals these practices. Examples of users enacting their identity via language can be found in Beers-Fägersten's (2008) investigation of hip hop interactive websites where users open their comments with *yo* and close it with *peace* to enact their in-group identity, whereas McDonald & Woodward-Kron (2016) finds that veteran members in support groups provide advice to new members by demonstrating that they live through a similar experience and establishing a shared identity. Examples of users' discourse construing their social world are shown in Hunt & Harvey (2015) and Sotillo & Wang-Gempp (2016). Collocates of the keywords of the main topics in the respective discussions, that is *anorexia* and *politicians* provide insights into users' experience and conceptualization of the topics.

For stance-taking in online spaces, modalized claims are often used, except in political discussions, for establishing dialogic relationship with others, even though the users may have different views (Drasovean & Tagg, 2015; Hewings et al., 2009; Sotillo & Wang-Gempp, 2016). It is interesting to note that while Hewings et al. (2009) show *do you think* or *what do you think* can be a way to open a dialogue, Sotillo & Wang-Gempp (2016) show that they can be used to challenge others, especially when users hold different ideologies. This shows that the same language pattern can be employed for different practices, and that a question format is not necessarily a question; that is, a form may have different functions depending on the context. Therefore, an in-depth qualitative analysis of discourse practices is needed.

However, for present purposes, there is a limitation common in these corpus studies. Although these studies reveal various discourse practices for challenging others or entertaining others' voices (Drasovean & Tagg, 2015; Sotillo & Wang-Gempp, 2016), there is no mention of user-user interactions in the threads; that is, which comments are challenged or entertained. These corpus studies seem to assume that the discourse is similar across different types of comments, such that there is no differentiation between initiating posts, independent posts or replies in the analyses.

Furthermore, how the discourse practices may trigger others' response in the threads also remain unexplored, probably because of the limitation of concordancing that usually does not go beyond the lines of text. Thus, this thesis will extend these corpus studies on online discourse by examining the linguistic features and discourse practices of initiating posts, independent posts and replies, while also employing micro-analysis of threads to further explore the discourse practices that move forward the conversations within a thread (see section 4.7).

It is worth noting that this thesis considers only corpus studies that utilize the established methods, keyword analysis, collocation analysis, concordancing, and involve both quantitative and qualitative analysis. It should be acknowledged that there are other corpus studies on online discourse utilizing methods such as type/token, lexical density (Riordan & Murray, 2010; Yates, 1996), and key semantic categories (Collins & Nerlich, 2015; Potts & Semino, 2017). The type/token and lexical density calculation are quantitative analyses that provide an indication of how varied or rich a text is in terms of lexical items. This information is not related to the aim of this thesis, so it is not considered here. The analysis of key semantic categories, that is semantic categories used significantly more often, requires annotating each word in the corpus prior to the analysis. Usually this semantic tagging is achieved automatically, for example by WMatrix (Rayson, 2008). However, as mentioned earlier, this thesis aims to analyse the actual language used, therefore, instead of codes or semantic categories, I will take the word forms, that is unannotated data, as the basic unit of quantitative analysis.

4.5 Concepts and methods in corpus linguistics

4.5.1 Corpus

A *corpus* is a sample of naturally occurring language collected for addressing research questions related to language usage, discourse or social practices. A corpus needs to be balanced and representative of the language usage or social phenomenon that a researcher examines (McEnery & Hardie, 2012). A corpus also needs to be large enough to facilitate the discovery of repeated

patterns which might not be visible if a smaller corpus were used, and to capture relatively rare yet important language patterns (Baker et al., 2008). Various corpora have been compiled by researchers to examine different types of language usage and social practices, for example, text-messaging (Tagg, 2012), online discussion in distance learning courses (Yates, 1996), business spoken language (Cheng, 2004), and media portrayal of refugees (Baker et al., 2008).

Large-scale corpora have also been compiled for examining general language usage, for example British National Corpus of spoken and written English (BNC, Leech et al., 2001; Love, Dembry, Hardie, Brezina, & McEnery, 2017). BNC is often used as reference corpus for keyword analysis, as will be discussed next. Two large-scale corpus studies describe comprehensively the form and function of American English (Biber et al., 1999) and British English (Carter & McCarthy, 2006), covering both spoken and written language. These two works will be referred to for the interpretation of keywords and linguistic features in this thesis because they cover most words and linguistic features in English.

4.5.2 Keyword Analysis

Keyword analysis is the main corpus method used in this thesis to explore the linguistic features and discourse practices of initiating posts, independent posts and replies. It is a corpus-driven, or data-driven approach that starts from quantitative analysis, then moves onto qualitative analysis for interpretation.

Quantitative analysis

Statistical analysis is conducted on every single word in a corpus (see section 4.5.3 for the statistical procedure). Typically, different forms of a word, such as *say*, *says* and *said*, are considered as separate words in keyword analysis, because they may have different discourse functions (Flowerdew, 2008; Grabowski, 2015; Holmes & Nesi, 2009; McEnery, 2016). A word is considered a *keyword* when its frequency in the corpus is found to be significantly higher than its frequency in a reference corpus. A reference corpus is typically a large-scale corpus of general language usage, such

as the BNC, as used in some corpus studies (e.g., Collins, 2019; McEnery, 2016). Given that the keywords are used significantly more often in the corpus than in the reference corpus, they could be indicative of the “aboutness” or style of the texts in the corpus (Baker, 2004). For example, O’Keeffe and Walsh (2012) find deitic *next* is one of the keywords in the classroom discourse corpus when compared to a general language corpus. It was used to signpost learning journey, as in *next week*. Based on the keywords found, they identified several core actions in classroom, such as demonstrating and sequencing, and feedback on elicitations. The keywords resulted from quantitative analysis and its interpretation by qualitative analysis reveals the subject matters or discourse practices in the corpus.

Keyword analysis can also be conducted by comparing sub-corpora within the corpus under examination, instead of comparing against another corpus. The sub-corpora comparison is useful for investigating within-corpus variation and has the advantage of not using an arbitrary reference corpus that might not be comparable to the corpus under examination (McEnery, 2016). The keyword analysis is a test of difference, and the reference corpus should thus only differ from the corpus under examination in terms of the factors that are relevant to the research purposes, rather than other unexplained confounds. One example of sub-corpora keyword analysis was conducted by Brookes & Baker (2017) who compare feedback to the NHS receiving rating 1 against feedback with other ratings to understand what makes a good NHS service. The sub-corpora comparison will also be used in this thesis to examine the different types of comments, i.e., initiating posts, independent posts and replies.

Qualitative analysis

There are generally two ways of conducting qualitative analysis in keyword analysis to interpret the keywords found. According to Baker (2004), one way is to categorize keywords based on their general functions in the corpus. I call it functional grouping in this thesis. The functional grouping is to group keywords based on their semantic meaning and communicative function in the context they occur, so as to provide an overview of the general trend of discourse practices in the

corpus. For example, O’Keeffe and Walsh (2012) categorize the three keywords *so*, *ok*, *alright* in their study as discourse markers. To get a glimpse of the general functions of each keyword, corpus methods such as concordance reading and collocation analysis (see section 4.5.4 and 4.5.5 for details) are typically used to investigate each keyword (Baker, 2004). It must be acknowledged that the functional grouping of the keywords is subjective and interpretative in nature. Furthermore, the functional group of a keyword only represent its most salient function and is not exhaustive of all instances of the keywords in the corpus (McEnery, 2016). In short, the functional grouping provides an overview of the ‘aboutness’ and style of the corpus based on all the keywords found.

Another way of conducting qualitative analysis of the keywords is to conduct an even more in-depth discourse analysis of selected few keywords. For example, O’Keeffe and Walsh (2012) further explore the keyword *if* and conclude three general functions for *if*-conditionals in their corpus – pedagogic illustration, projecting and demonstrating. They also conduct micro-analysis of a few episodes of conversation to illustrate how some keywords are used in different types of classroom talks. In another corpus study, as reviewed in Chapter 3, Wagner & Herbel-Eisenmann (2008) focus on only one keyword *just* and conclude that when teachers use it to instruct students to tackle a maths question in a particular way, it contributes to a reduced sense of agency among students. The in-depth discourse analysis is important for understanding actual language use and social practices given that interaction must be contextualized in its social context. Both types of qualitative analysis– functional grouping and discourse analysis– are conducted in this thesis to interpret and explore the linguistic features and discourse practices in initiating posts, independent posts and replies in the online discussions.

One advantage of keyword analysis is its bottom-up approach, such that all the words in the corpus are subjected to the statistical analysis. This could remove researchers’ bias, and allows for exploration of linguistic features which might not have been considered in previous studies or hypothesized to be relevant (Baker, 2004). This data-driven component of keyword analysis thus has two major functions– it reveals the general language patterns in the corpus and provides a point of

entry into the corpus, that is backed by statistical analysis, rather than decided by researchers a priori. Against the general language patterns revealed, the point of entry not only narrows down the amount of language data a researcher faces when conducting in-depth discourse analysis, but also supports the choice of discourse practices that a researcher focuses on (e.g., Wagner & Herbel-Eisenmann, 2008). However, it is worth noting that, although the statistical analysis of keyword analysis offers a degree of objectivity, the interpretation of the keywords is subjected to a researcher's interpretation and the theoretical or methodological framework of the qualitative analysis chosen.

4.5.3 Statistical analysis for keyword analysis

A keyword is a word used significantly more often in the corpus under examination than would be expected, given the word distributions in the corpus and the reference corpus. To put it in general statistical terminology, for every single word, a contingency table analysis is conducted such that the frequency of the word in the two corpora and total frequency of all words in the two corpora are subjected to statistical test (Baroni & Evert, 2016). The statistical test used in this thesis is log-likelihood ratio test, which has the benefit of not being biased by huge sample size differences between the two comparison corpora. In the present analysis, a word is considered a keyword when the p -value for the log-likelihood ratio test is $< .000000000001$. This strict p -value is adopted because the common practice of $p < .000001$ (Grabowski, 2015; Holmes & Nesi, 2009) is considered too liberal because the large number of comparisons conducted in this study could inflate type I error, that is rejecting a true null hypothesis. Translating it to keyword analysis, it means identifying a word as a keyword when in reality it is used similarly often in the language represented by the two corpora while the significance finding is due to chance arising from sampling. This strict p -value was also used by Flowerdew (2008) to obtain a reasonable number of keywords that are not too many or too few for in-depth qualitative analysis. In addition, following McEnery (2016), the normalized frequency of a keyword must be 5 per 100,000 words to ensure that the keyword is a common word

in the corpus. Sometimes, the keyword analysis will result in large number of keywords, so most research examine preliminarily the top 50 keywords (e.g., Grabowski, 2015) or top 100 keywords (e.g., Malavasi & Mazzi, 2010; Potts, Simm, Whittle, & Unger, 2014) and then in details a selected group of keywords.

Besides these criteria, this thesis also imposes another criterion for the keyword analysis – dispersion measure. It measures how equally distributed a word is across all the texts within a corpus. A keyword can be high in frequency but is only used frequently in a particular text and not used in other texts within the corpus. For example, a subject-related word, *palliative* is used frequently in a palliative care MOOC in the corpus but is not used at all in other MOOCs. Such a keyword may not be a true keyword, given that keywords are supposed to be common across the corpus (Baker, 2004), and I am interested in non-course specific discourse, unlike Collins (2019) who investigates users' use of the technical term *face* in his MOOC. Several dispersion measures have been used in corpus linguistics, including the commonly used Juilland's D (see Gries, 2008 for a review). However, in the present analysis, Gries' Deviation of Proportion (DP, Gries, 2008; Lijffijt & Gries, 2012) is used instead because it accounts for the different sample sizes of the different MOOCs comprising the corpus.

DP "can theoretically range from approximately 0 to 1, where values close to 0 indicate that a [word] is distributed across the n corpus parts as one would expect given the sizes of the n corpus parts. By contrast, values close to 1 indicate that a [word] is distributed across the n corpus parts exactly the opposite way one would expect given the sizes of the n corpus parts" (Lijffijt & Gries, 2012, p.147). Translating this to this thesis, the corpus parts are the MOOCs in the corpus. The DP is calculated to examine if a keyword found is distributed evenly across the 12 MOOCs in the corpus such that it could be established as a common linguistic feature in the corpus. Table 4.3 illustrates the distribution of four words which have different DPs. They are potential keywords of initiating posts based on the log-likelihood ratio test (to be introduced in section 4.6 and Chapter 6). *Pension* which DP is 0.93 is only used in three out of 12 MOOCs and the majority occurs in finance-1 MOOC.

Women which DP is 0.54 is used in most MOOCs although most occurs in ancient-1 and corpus-1. In contrast, both *tried* and *say* which DP is less than 0.30 are used in all MOOCs, although the frequency of *say* with a DP of 0.08 is distributed more evenly across all MOOCs. Therefore, although all four words can be keywords for initiating posts according to the log-likelihood ratio tests, only *tried* and *say* whose DP is below 0.3 are considered keywords.

Table 4.3 The distribution of four words with different DP.

MOOCs ¹	Total word count of initiating posts ²	pension (DP=0.93)	women (DP=0.54)	tried (DP=0.28)	say (DP=0.08)
		Normalized Frequency ³	Normalized Frequency	Normalized Frequency	Normalized Frequency
accessibility-2	92680	0	3	46	47
ancient-1	330828	2	113	15	53
code-1	105238	0	0	112	32
corpus-1	500678	0	40	45	74
dyslexia-1	406832	0	1	30	68
finance-1	188831	380	17	7	60
management-4	73449	1	1	12	49
moons-1	201854	0	1	12	60
nutrition-4	197137	0	14	25	71
oceans-1	92151	0	3	26	60
palliative-1	70113	0	7	7	60
soils-1	140227	0	1	9	46

¹The details of the MOOCs and the corpus to be examined will be introduced in Chapter 5.

²The normalized frequency is based on the word frequency of initiating posts because this example of DP is from the keyword analysis of initiating posts (to be introduced in section 4.6 and Chapter 6).

³The normalized frequency is per 100,000 words, based on the total word count of initiating posts in each MOOC.

Previous research which used Juilland's D has a cut-off of 0.8 (The scale of Juilland's D is opposite to DP), but this is an arbitrary cut-off subjected to researchers' discretion (Paquot & Bestgen, 2009). If a similar cut-off for DP is taken, i.e., 0.2, this will result in fewer than 50 keywords, fewer than the common practices of keyword analysis as reviewed earlier. Therefore, a more liberal yet still conservative 0.3 cut-off is adopted in this thesis.

Keyword analysis is a test of difference in word frequency between two corpora. Therefore, a measure of this difference, i.e., effect size, also needs to be calculated to better understand how

often a keyword is used in the corpus under examination compared to the reference corpus. Various effect size measures have been proposed in corpus linguistics, for example ratio, odds ratio and log ratio (Gabrielatos, 2018). In this thesis, relative risk, that is the ratio of the normalized frequency of a word in the two corpora, is used to report the effect size of the keyword analysis. Relative risk is one of the effect size measures used in statistics for log-likelihood ratio test of the contingency table, on which keyword analysis is based (Baroni & Evert, 2016). It is easily interpretable because it effectively reveals how many times a keyword is used more often in one corpus compared to another. While there is no consensus yet among corpus linguists regarding the cut-off threshold of effect size for a keyword (Gabrielatos, 2018), this thesis only reports effect size to illustrate the difference in the frequency of a keyword in the two corpora, rather than as another criteria to decide whether a word is a keyword. This decision is also in line with APA guidelines on reporting effect size alongside statistical significance (*Publication Manual of the American Psychological Association*, 2020).

4.5.4 Collocation analysis

Collocation refers to the observation that certain words tend to co-occur frequently in the corpus.

The collocates of a word contributes to its meaning and function, and may unveil the context and discourse surrounding the word. For example, Gabrielatos & Baker (2008) found collocates of *refugees* and *asylum seekers* in news media include *flooding*, *pouring*, which construe a negative metaphor towards them. In collocation analysis, two words could be labelled as collocating with each other side by side or within a window of n-words, depending on the research purposes.

Following Gabrielatos & Baker (2008), this thesis investigates collocates in the 5-word window on either side of a word. Additionally, a statistical measure, mutual information 3 (MI^3) is used to decide on the importance of collocates of each word. The MI^3 is an effect size measure rather than a test of significance. It measures how much the observed co-occurrence frequency of the two words exceed

expected frequency (Evert, 2008). In this thesis, collocation analysis is used along with concordancing to facilitate interpretation of the keywords for their functional grouping.

4.5.5 Concordance

Concordancing is the main method for qualitative analysis in corpus linguistics (McEnery & Hardie, 2012). Concordance lines show the word of interest in their co-text, i.e., a span of characters or words, in a vertical format, as shown in Figure 4.1. The usage and senses of the word of interest across the corpus could thus be analysed qualitatively (Sinclair, 2003). Typically, corpus tools allow alphabetical sorting of concordance lines based on words on the *n*-th position on the right or left of the word of interest such that recurrent patterns can be revealed to facilitate analysis. The number of times of recurrent patterns occur can typically be counted, and some researchers call them clusters or lexical bundles (Beers-Fägersten, 2008; Biber et al., 1999). The word of interest could be a keyword or word that is related to a social phenomenon or discourse practices which researchers are interested. In this thesis, keywords of initiating posts, independent posts and replies will be concordanced to examine the general patterns of each keyword to decide on their functional grouping.

Figure 4.1 Concordance of the word “FutureLearn” in the corpus.

icipate in competitions and sports Hello , I love	<FutureLearn>	courses and have taken quite a few of them . Alth
am retired person and love the chances afforded by	<FutureLearn>	. I am interested to see if the issue of disabili
s of the modern world . I have studied a number of	<FutureLearn>	courses and apart from keeping my own mind active
a very curious but lazy individual but find these	<FutureLearn>	courses wonderful . As a family historian I would
course in helping me understand that ! There is a	<FutureLearn>	course on wellbeing , and it could n't define it
d the digs at Vindolanda and to take the Newcastle	<FutureLearn>	course on Hadrian 's Wall . The digs were outstan
istory that I would not have landed on myself . At	<FutureLearn>	I have taken the regular history courses like Had
e graduating , I 've studied Roman Britain and the	<FutureLearn>	Portus course . Nearing retirement I 'm keen to f
History and have enjoyed doing other history based	<FutureLearn>	courses . I have worked in the health system in A
, volunteering , genealogy , and taking courses on	<FutureLearn>	. I saw a Pompeii exhibit in Seattle a year ago a
ithin heritage museum environment . I enjoy taking	<FutureLearn>	courses as it keeps my mind active learning new t
cepts . Completely fascinating . This is my second	<FutureLearn>	course after Hadrian 's Wall last fall . I 'm a f
ls ! Good Evening ! Hello , all , this is my first	<FutureLearn>	course ! I 've had an interest in Greece and Rome
ly. .) would love to learn more . Hi Theoni . The	<FutureLearn>	course on England in the time of Richard III ('la
course . Welcome , Bryony . I personally like the	<FutureLearn>	model - short ' steps ' and sharing of resources
s too expensive now , but I 'm really grateful for	<FutureLearn>	, especially courses like this that send me off i
latest so far . I was really bussy following other	<FutureLearn>	courses . Ready to follow the course and comments
e been ... Yes Helen , thank you so much . I enjoy	<FutureLearn>	so much . I know this will be very interesting .
y and thank you again Alice . Very good Yoko . The	<FutureLearn>	site also brings me joy as well . Yoko . It is a
portant to stay positive , keep the brain active (<FutureLearn>	courses !) and get outside as often as possible
hocolate " , describes me to a T. I also walk , do	<FutureLearn>	, and also as I am retired , am an avid USA membe
reasons . Oh and I keep my brain going with doing	<FutureLearn>	courses , and inventing things , though I always
sit ? (I do recall from previous conversations on	<FutureLearn>	that my mother lives in your approximate area - s
join in or continue , if they do n't want . But on	<FutureLearn>	we have discussion forums so that people can disc
ith our bitter sweet memories . I recently did the	<FutureLearn>	English Literature and Menatal Health course whic
I have hives and allergies . And an obsession with	<FutureLearn>	courses . This course is the best yet . I bet I f
an additional , even more baffling entity . On the	<FutureLearn>	Literature and Mental Health course there was ref
opathy We 've been discussing Takotsubo on another	<FutureLearn>	course . It is very rare and although it was orig

4.6 Applying corpus linguistics to online discussions in FutureLearn

The keyword analysis is the starting point of this thesis, and sub-corpora comparison is used. To reveal the keywords of initiating posts (called *initiating keywords* afterwards) and the keywords of independent posts (called *independent keywords* afterwards), these two types of posts are compared with each other in the keyword analysis. To reveal the keywords of replies (called *reply keywords* afterwards), replies are compared with both initiating posts and independent posts.

The sub-corpora comparison is used for two reasons. Firstly, this thesis aims at examining the linguistic features and discourse practices characterizing each type of comments, rather than characterizing the comments in the online discussion as a whole. This is unlike previous studies (e.g., Tagg, 2012; Yates, 1996) which compared the language of digitally or internet mediated communication to spoken and/or written language to examine how it differs and resembles language of other modalities. Secondly, each type of comments makes a good comparison for each other, because the only difference among them is their nature within the discussion space itself (Brookes & Baker, 2017; Ksiazek & Lessard, 2016). Initiating posts and independent posts are both new posts except the former receive replies, and the latter do not. Replies differ from both initiating posts and independent posts because replies are responses to others' comments. Therefore, the sub-corpora comparison will only reveal the differences between these three types of comments, instead of other issues not relevant to the current research purpose, for example differences in genre or platform.

To address the research questions, qualitative analysis is then conducted on the three types of keywords resulting from the comparisons, i.e., initiating keywords, independent keywords and reply keywords. Each keyword will be functionally grouped based on concordance reading and collocation analysis. The functional grouping thus characterizes each type of comment in terms of linguistic features and discourse practices realized by the keywords. Based on the keyword analysis, how users initiate, sustain or hinder dialogic conversations in the online discussions can thus be

identified. The keywords will be italicized in the description of the findings and in the example comments presented in this thesis.

It is worth noting that the functional grouping of initiating keywords, independent keywords, reply keywords, as presented in Chapter 6, 7 and 8, does not follow a strict exact coding and counting procedure, such that the number of times a keyword is used for a specific function is not counted. Firstly, the big data set does not allow such an intensive procedure, unless I choose to focus only on a few words, like Wagner & Herbel-Eisenmann (2008) who only examine *just* in their classroom corpus. Secondly, and more importantly, it is not my purpose to quantify the discourse for generalization, but to interpret discourse based on demonstrable language data. Although the interpretation is subjective, the functional grouping and analysis is based on extensive reading of concordance lines and threads, collocation analysis that reveal recurrent patterns. Furthermore, the keywords found, as well as collocation analysis, is driven by word frequency information and statistical analysis, therefore the findings are supported by quantitative data.

In this thesis, the interpretation of the keywords is primarily directed by the theoretical concepts of dialogic space (Martin & White, 2005) and intersubjectivity (Du Bois, 2007), while past research on discourse related to the keywords is also consulted, such as *if*-conditionals (G. Ferguson, 2001). As mentioned earlier, the comprehensive grammar of American and British English by Biber et al. (1999) and Carter & McCarthy (2006) will also be referred to. Meanwhile, the educational context and FutureLearn platform design which might influence users' discourse practices in this online space will also be taken into account because user-user interactions and user-content interactions are mediated by the technology and design of websites and platforms (Herring, 2004; 2013). It is worth noting that, in a preliminary analysis, an automatic semantic tagging with WMatrix (Rayson, 2008) was conducted but I decided not to proceed with it. This was because the categories in WMatrix are too broad for the current analysis of discourse practices. Furthermore, as mentioned earlier, this thesis aims to analyse the actual language used, so the semantic categories determined by WMatrix are not considered.

4.7 Micro-analysis of discussion threads

The fact that an initiating post and replies form a thread should not be overlooked. The micro-analysis of threads extends the corpus analysis by situating the linguistic features and discourse practices into their context in the threads to explore how they are employed to establish and sustain a dialogic conversation. The micro-analysis of threads is informed by Giles et al. (2014) who adapt the principles of (spoken) conversation analysis (CA) to examine online data. The principles of CA in spoken conversation will be briefly explained and how they are adapted and applied in online discourse will be explained in the following subsections.

Numerous studies, including some that have been reviewed earlier (Baym, 1996; Paulus, 2006; Stommel & Koole, 2010; Ziegler et al., 2014), have employed this kind of approach to investigate conversations in text-based asynchronous online communication. The micro-analysis of threads in this thesis focuses mainly on how the reply keywords realize discourse practices in the turn-taking among users, especially how they are received by others and co-constructed by users across the turns. The analysis of the threads will also be directed by theoretical concepts of dialogic space and intersubjectivity, rather than concepts in CA, although intersubjectivity is also utilized in CA (Schegloff, 1992) and the membership categorization analysis in CA is similar to identity performance.

4.7.1 Turn-taking in CA

CA investigates social interactions by examining speakers' discourse practices in their turn-taking (Sacks, Schegloff, & Jefferson, 1974). A conversation can be seen as a sequence of actions or conversational moves, such that each turn can be considered as immediate response to the preceding turn, and at the same time be designed to elicit response in the next turn. Analysis of adjacent pairs, i.e., the turns in sequence and in connection with each other, is important for understanding social interactions such as question and answer, repairing, opening routines, leave-taking, stance-taking (e.g., Pomerantz, 1984; Schegloff, 1992). Breakdown of the routine of the turn-

taking sequence allows the interlocutors to recognize breakdown of intersubjectivity, in turn repair their talk (Schegloff, 1992). Furthermore, the epistemic status of interlocutors can be co-constructed turn-by-turn via the expression of epistemic stance among interlocutors (Heritage, 2012). Whether a turn conveys or requests information can be determined by the epistemic status among the interlocutors.

However, in text-based asynchronous online discussions, turn-taking of the same conversation might not be adjacent given its polylogal nature (Baym, 1996; Herring, 2001, 2004). It is not subjected to the constraint of face-to-face conversation, such that some turns can be left unresponded to and users can leave a conversation without any leave-taking. Apparently, the delay and latching between turns, pauses, or fillers which are the discourse devices in spoken conversations are not available in asynchronous online discussions. Therefore, strictly speaking, the sequential analysis of turn-taking in CA is not applicable in online spaces. However, Gibson (2009) suggests that the “adjacent pairs”, although disrupted by other turns, can be identified and analysed accordingly.

Therefore, the turn-taking can be loosely operationalized as corresponding to the initiating post and all the replies within the thread in online spaces. Understanding the routine of turn-taking in online spaces is crucial for one to be an effective member in online groups to engage in discussions with others. As reviewed earlier, Baym (1996) reveals how agreement and disagreement are used in response to previous messages while also creating social relationships or moving forward a conversation. Kääntä & Lehtinen (2016) find adjacency turns consisted of first story and second story, typically coupled with implicit or explicit agreement, accomplish alignment and affiliation with others. In contrast, Stommel & Koole (2010) show that mismatch of adjacent pairs between new members and veteran members may result in new members being not welcomed. This happens when the new member engages in problem telling in their opening post (e.g., “I had to tell someone”) and do not acknowledge the advice given by veteran members who treats the opening post as advice seeking. Stommel & Koole’s (2010) findings also suggest that online support group is

not necessarily welcoming to all members, but those who display discourse practices that are in line with the communicative norms, in this case the turn-taking norms.

This thesis does not examine the sequential organization of turn-taking in online discussions. However, I will follow Gibson's (2009) suggestion to identify the "adjacent pairs" in the thread. Then, I interpret linguistic features and discourse practices in each turn in relation to the turn(s) it responds to and the turn(s) it triggers. This way, the interpretation of the discourse practices is based on how turns are received by others as evidenced by the responses, rather than purely based on my interpretation. Specifically, in Chapter 8, the function of *agree to disagree* is examined by exploring the turn prior to and following it, whereas in Chapter 9, the function of the URLs posted in a turn is construed by the turn it responds to and the responses it receives.

4.7.2 Context in CA

Strictly speaking in CA, researchers only analyse and draw interpretation based on what is spoken, alongside pauses and sounds made in an episode of conversation. The context where a conversation takes place and characteristics of interlocutors are only taken into account when speakers make them relevant in their talk (Antaki & Ardévol, 2005). It can be argued that this assumption is also valid in most asynchronous online discussions because users do not know each other in person, but construe their identity based on how they write and what they mention through their interactions, and perhaps their online profiles.

One relevant analysis in CA, membership categorization analysis, investigates how interlocutors enact their social identity in the local context of their conversation. Stommel & Koole (2010) apply this analysis to examine how users orient to their membership categories via their discourse practices in an online support group. They find that new members usually use categorization, for example different diagnoses of eating disorders, to legitimate their participation in the group. This self-presentation is typically reciprocated by veteran members who mention their past experience. In contrast, a new member who "glamorizes" their diagnosis enact themselves as

an out-group member, and in turn receives advice from veteran members asking them to leave pro-anorexia as a membership category behind. This finding reveals the way users employ categorization to enact an identity in social interaction online, and how this is recognized and acted on by others. As reviewed in Chapter 3, users enact various aspects of their identity in online spaces to legitimate their claims, align with those similar to them, or shift focus of the conversations (Grabill & Pigg, 2012; Vásquez, 2018).

Although I will not conduct membership categorization analysis in this thesis, I will follow the rationale that context and identity will only be considered when users make them relevant in their turns. The larger context of the digital environment, in this case FutureLearn MOOCs will still be taken into account in my analysis, as recommended by Giles et al. (2014).

4.7.3 Procedures in CA

CA investigates social practices turn-by-turn. Such a fine-grained analysis does not allow examination and presentation of a large number of episodes of conversations, but only selected conversations or part of them. Although some might argue that this kind of analysis is idiosyncratic and not generalizable, researchers do follow a systematic procedure in CA, and it could safely be assumed that the findings highlight what can possibly be achieved by different discourse practices (Heritage, 2004).

According to Heritage (2004), researchers in CA can first identify some candidate practices which are distinct and relevant to their research interest, by preliminary reading of the data. Then, a collection of conversations that involve these practices are compiled. To analyse the data following CA principles including sequential analysis of turn-taking, relevant turns are located in each conversation and analysed. Finally, the researchers may narrow down their choices of candidate practices to the practices they systematically find in the collection of conversations. It is not necessary that the social practices found are general across the collection of conversations. Rather, CA pays attention to practices that have a “distinctive character” and are “distinctive in [their]

consequences for the nature or the meaning of the action that the turn implements” (Heritage, 2004, p. 228). Furthermore, a breakdown of turn-taking sequence also informs understanding of routine turn-taking, and how interlocutors repair the conversation (Sacks et al., 1974).

The micro-analysis in this thesis will loosely follow such a procedure. Admittedly, the large number of discussion threads in my corpus, 32,334 threads (see Chapter 5), does not allow me to do a preliminary reading of all threads. It is done instead by my extensive reading of discussion threads in the corpus when examining initiating keywords and reply keywords, and other selective reading of discussion threads including:

- the ten longest threads in each MOOC involving users only,
- the ten longest threads in each MOOC involving facilitators and users,
- all threads involving at least two users who repeatedly contribute, and
- threads containing the greatest number of keywords.

The discourse practices to be analysed in the micro-analysis are thus informed by my preliminary reading and analyses of some of the threads, as well as driven by the linguistic features and discourse practices found in the keyword analysis. The justification of the discourse practices to be focused on will be presented in Chapter 8 where micro-analyses are conducted. The selection of threads and discourse practices is also partly directed by the literature and theoretical concepts of dialogic space and intersubjectivity.

Specifically, for the investigation of *agree to disagree* in Chapter 8 and examination of how users conceptualize Wikipedia in online discussions in Chapter 9, all the threads containing the related phrases are investigated. Selected threads are then presented for the micro-analysis. In Chapter 9, the micro-analysis of the URL-posting practices starts from a preliminary reading of threads containing the highest numbers of URLs posted or the mention of the reply keyword *link* or *links* that leads to the identification of the distinctive practice of employing URLs for stance-taking. The analysis process will be explained in Chapter 9.

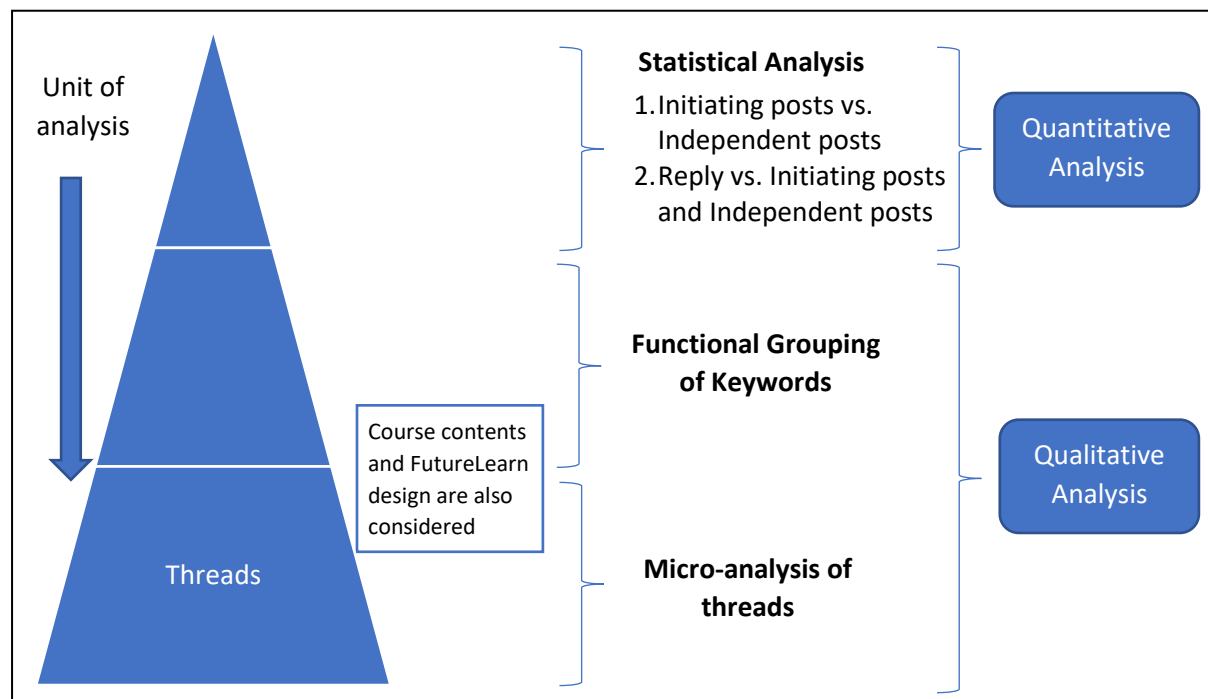
In short, although online interactions are different from spoken conversations, following Giles et al. (2014) as well as other researchers who have adopted CA to investigate online conversations (Baym, 1996; Paulus, 2006; Stommel & Koole, 2010; Ziegler et al., 2014), I draw on different principles and procedures in CA to conduct micro-analysis to explore dialogic conversations among users. This micro-level analysis illustrates how users respond to each other by drawing on linguistic features and discourse practices, thus extending the findings of corpus analysis which at times reveals only discourse practices at a broader level rather than at the local level of a thread.

4.8 Conclusion

A mixed methodology of corpus linguistics is adopted in this thesis, complemented by micro-analysis of threads, to explore how users employ their language to initiate and engage in conversations with others. This decision comes from the assumption of language-in-action, which is attested by numerous studies reviewed thus far revealing how users' discourse practices can establish relationships, enact identity and bring about actions from others. Neither methodology has been fully utilized in MOOC research which typically assumes language as reflecting thinking and reduces users' textual contribution to codes for counting purposes. Therefore, this thesis will not only shed light on users' discourse practices in online spaces, but also introduce a new methodological perspective to MOOC research.

Overall, the analysis in this thesis is a bottom up approach that starts from the word level then expands to the discussion threads by utilizing both corpus linguistics and micro-analysis of threads, as visualized in Figure 4.2.

Figure 4.2 Summary of methodology in this thesis.



Note. Unit of analysis expands from single word to threads.

It starts from a keyword analysis characterising the general patterns of each type of comment; that is initiating posts, independent posts and replies. This extends previous corpus studies that do not make such a differentiation and allows for a systematic analysis of the different types of comments in the online discussions. This differentiation is crucial for understanding the dialogic nature of the online discourse because initiating posts and replies are where explicit user-user interactions occur, whereas independent posts are not involved in any conversation thread.

The keyword analysis considers every single word in the corpus, such that the analysis is driven by the data, rather than decided a priori by researchers. The quantitative analysis is based on formal structures of language, unlike the coding and counting paradigm in which the quantitative analysis is conducted on codes which are subjected to pre-defined categories and human interpretation. Importantly, qualitative analysis of keywords based on their functional grouping and micro-analysis of threads not only provides the interpretation for the quantitative analysis but also facilitates our understanding of the linguistic features and discourse practices of user-user

interactions in the online discussions. This analysis thus extends previous studies that only investigate linguistic features. In short, the findings presented in Chapter 6 to 9 will reveal both the general patterns and detailed analysis of discourse practices in initiating posts, independent posts, replies and of URL-posting.

Chapter 5

The FutureLearn Corpus

This chapter documents the creation of the FutureLearn corpus (FL corpus) and the technical procedure of annotating the metadata, storing and querying the corpus. The technical procedure of conducting the statistical analysis and qualitative analysis is also explained. Ethical consideration in using the users' textual contribution in online spaces for research purposes is discussed in light of the public-private nature of online discussions in MOOCs. This chapter ends by describing the FL corpus in detail.

5.1 Introduction: The FL corpus

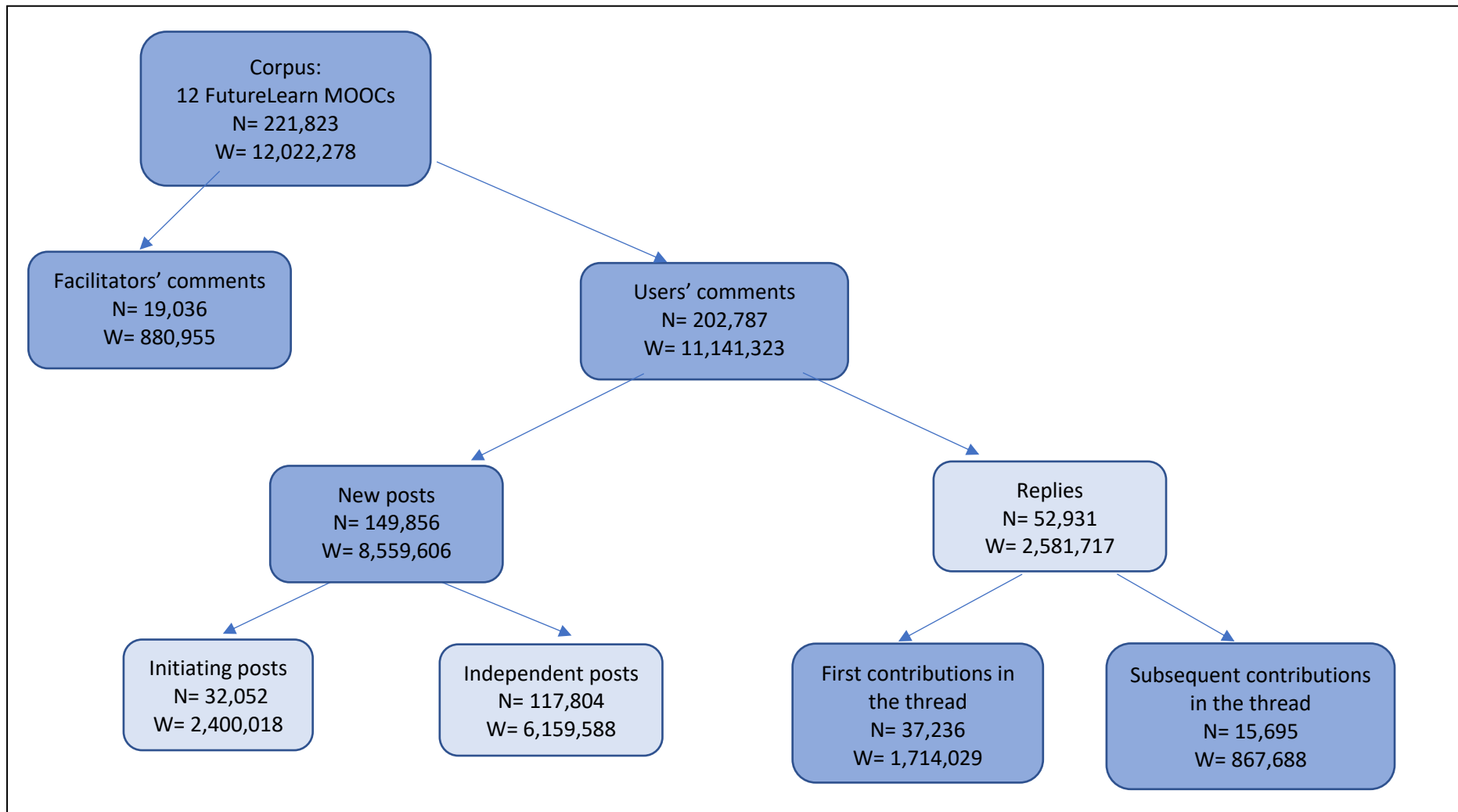
The FL corpus consists of all the 221823 comments in the online discussion of one presentation of each of 12 FutureLearn MOOCs, with a total wordcount⁵ of 120,22278, all in English. Among all the comments, 202787 were contributed by 22866⁶ learners, 19036 were contributed by 104 facilitators. This thesis is concerned primarily with learner discourse, and facilitators' comments are not analysed. This is in part because preliminary analysis suggests that learners' comments may mirror general online discourse more than facilitators, who mainly respond to learners' queries or pre-empt course content. Facilitators' comments will be referred to whenever necessary. The corpus information regarding facilitators' comments in the 12 MOOCs can be found in Appendix A as they remain part of the FL corpus. From now onwards, I describe the learners as *users* in this thesis because I am investigating their interactions from the point of view that online discussions can

⁵ It also includes punctuation, as the tokenization program – Treetagger (Schmid, 1994) used in this thesis considers them as such. Although token is generally used in corpus linguistics, I use wordcount instead because type and token issue is not related to the aims of this thesis.

⁶ 629 of them participating in the discussions of more than one MOOC.

happen in any general online space, rather than from the learning perspective. The corpus can be further broken down into sub-corpora based on the nature of the comments, as illustrated in Figure 5.1. The three sub-corpora initiating posts, independent posts and replies are compared in the keyword analysis, although the replies can be further categorized based on whether a user has contributed in a thread before. Further description of the corpus will be presented after the compilation of the corpus is explained.

Figure 5.1 Components of the corpus compiled and examined in this thesis.



Note. The components in lighter colour are the sub-corpora for keyword analysis. N refers to the number of comments, W refers to wordcount.

5.2 Corpus compilation

5.2.1 Data collection

The FL corpus consists of all the comments contributed in the discussion space of 12 FutureLearn MOOCs. The sampling of these 12 MOOCs is a convenient yet purposeful decision. I decided to collect data from MOOCs of different disciplines and length and offered by different universities to ensure that the corpus is not skewed towards any of the variables. The details of each MOOCs are described in section 5.4.1. I approached MOOC managers of five universities whom I got to know through conferences and FutureLearn Academic Network meeting, and applied for access to the data and the MOOCs following the guidance of each university, as summarized in Table 5.1. One university did not approve my request. This is a convenience sampling because I only gathered data from MOOC managers I know personally, and the MOOCs selected were those available at the time or recommended by the MOOC managers. At the same time, I ensured that all the MOOCs were run with the same platform features because Futurelearn has been changing their design and business model across the years, as summarized in Table 5.2.

Table 5.1 The data access process with different universities

	Number of MOOCs included in the corpus	Application procedure	Outcome
The Open University	4	The Open University Human Research Ethics Committee's approval	The MOOC manager provided me data and granted me access to the MOOC on the FutureLearn as a learner.
Lancaster University	4	The Open University Human Research Ethics Committee's approval and email approval from individual course leaders.	Four out of five course leaders replied and approved. The MOOC manager provided me data and granted me access to the MOOC on the FutureLearn as a learner.
University of Southampton	3	The Open University Human Research Ethics Committee's approval and approval from their MOOC project leader.	The MOOC manager only granted me access to their own data storage platform to download the data needed. I chose the three MOOCs that I happened to be a learner but never participated in the discussion, so that I would have access to the MOOC on the FutureLearn.
University of Aberdeen	1	The Open University Human Research Ethics Committee's approval and email approval from individual MOOC leader.	The MOOC manager provided me data and granted me access to the MOOC on the FutureLearn as a learner. They provided me data of the fourth presentation of the Nutrition and Wellbeing course because the data of previous presentations of the course is too large to send over.

Table 5.2 Changes in the features of FutureLearn since the data collection for the corpus.

	MOOCs in the corpus	Current MOOCs, as of 2020
Fee	Users can access the MOOC free of charge for as long as they want.	Users can access the MOOC free of charge during the run of the course and up to two weeks after it ends. They have to pay to unlock assessment and to have permanent access to the MOOC.
Replies	All the replies within a thread are shown by default.	Only one reply within a thread is shown by default but the number of other replies is given, and users can click "view" to see them.
New features on the platform since I collected my data	None	<ul style="list-style-type: none"> • Users can bookmark others' comments. • When users click reply under a post or a reply, the contributing user of that post or reply will be tagged with @ in the reply, although users can delete them. • Facilitators can pin comments to be always shown on the top. • Users can mute notification when others reply within a thread that they have contributed before.
Facilitators	Academia from the universities.	Some MOOCs recruit users who have taken the course before as facilitators.
Additional features designed by individual universities	None	Some MOOCs have trialed features such as discussion groups and word cloud summarizing comments posted.

5.2.2 Data processing

The data of the 12 MOOCs were processed before it could be encoded into the corpus tool and subjected to analysis. For each MOOC, three csv. files were provided from the MOOC managers. They were the standard output provided by FutureLearn for each university to download through their partner platform. One file contained the comments and metadata regarding the comments, as shown in Table 5.3. These were all the data needed for the corpus analysis in this thesis, and will be further explained below. The other two files were about activity and enrolments. The activity data recorded the timestamp of each user visiting a step and marking a step as “completed”, and was used to identify users who visit at least one step of the course. The enrolment data recorded users’

information such as gender, age, country, education level and employment status, but very few users provided such information, so these data were not included in the corpus and analysis. In some courses, the enrolment data also contained information about which user id belonged to facilitators, although not all facilitators were marked. Therefore, I visited each MOOC on FutureLearn and checked the introductory step where facilitators were introduced, and browsed through the discussion space in the first week of the courses because some facilitators only identified themselves in the discussion space. I gathered comments contributed by them, and mapped them back to the comment data file to identify their id accordingly. This identification is needed to differentiate facilitators' comments from learners' comments.

Table 5.3 Data in the comment file that is to be processed and encoded into the corpus

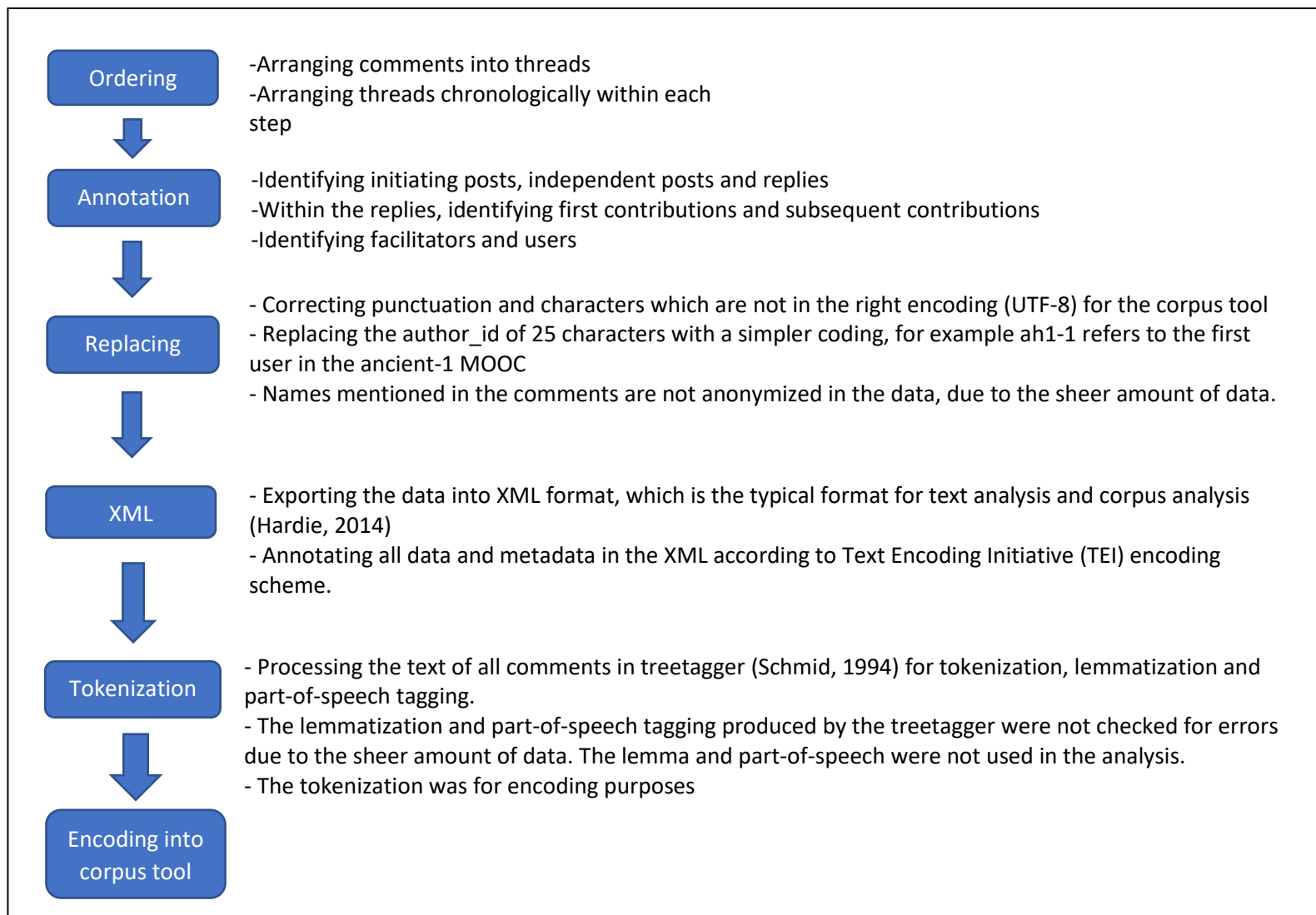
Data	Explanation
id	The id of the comment.
author_id	The user's id.
parent_id	The id of the post that the comment posted under. If there is a parent_id, the comment is a reply, if it is blank, it is either an initiating post or independent post.
step	The step where the comment is posted.
week_number	The week of the course where the comment is posted.
step_number	The step in the week where the comment is posted.
text	The comment.
timestamp	The date and time when the comment is posted. ⁷
likes	The number of likes the comment receives.

The comment data file contained the data, i.e., the comments, and metadata, i.e., data about the comments, to be processed and encoded into the corpus tool. The comment data file was arranged by the timestamp such that comments from different threads and steps were jumbled together, since the online discussion was asynchronous. As shown in Table 5.3, the data provided did

⁷ All universities provided the data in the same format as designed by FutureLearn, except University of Southampton. The university provided data in their own format, such that date instead of timestamp of posting is used, but it does not affect the corpus construction and analysis.

not include information regarding the types of the comments, i.e., initiating posts, independent posts, replies, and whether they were contributed by facilitators. Based on the id, parent_id, author_id and timestamp, I was able to categorize the comments into the three types of comments, and arranged the initiating posts and independent posts, and the replies within the threads according to the time of posting. Based on the author_id, two types of replies can be further identified: (1) *first contributions* in the thread, i.e., the reply is the first time a user contributes to a thread; (2) *subsequent contributions* in the thread, i.e., the reply is contributed by a user who has joined the thread before, either by replying or initiating the thread. Other pre-processing of the data is illustrated in . The processing was achieved mainly by using R programming (R core Team, 2018). This process was needed not only for encoding data into corpus tool but also to reconstruct the online discussion from the data file for qualitative analysis. The processed data was stored in an R dataframe object as a database.

Figure 5.2 Data-processing



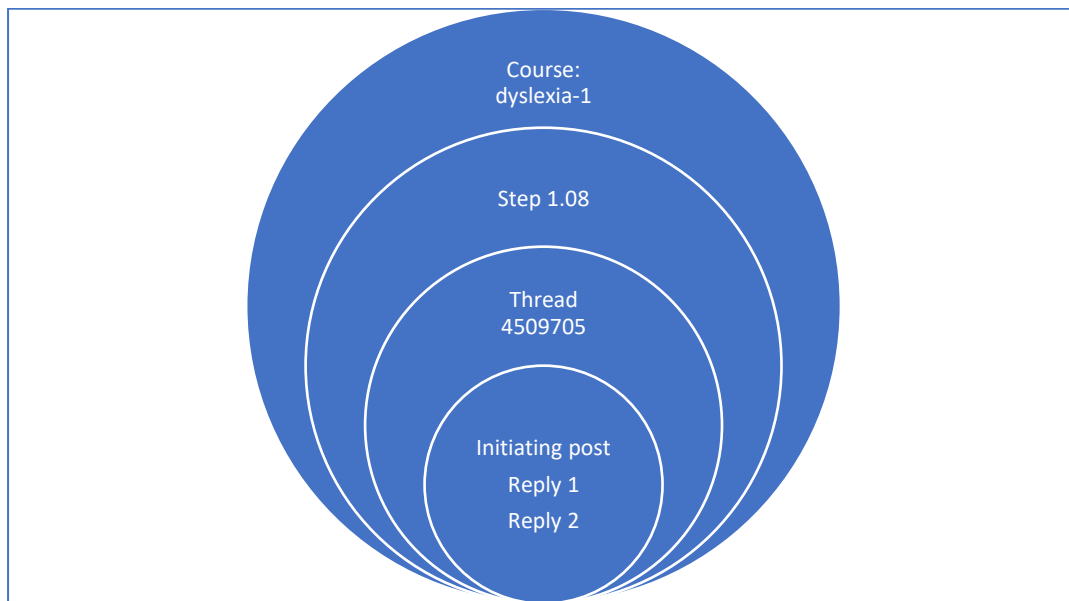
5.2.3 Corpus tool

The data and metadata were encoded into the Corpus Workbench tool (CWB, Evert & Hardie, 2011).

The CWB tool is a backend corpus tool with efficient data structure and retrieval for concordancing as well as frequency counts and sorting. It is the backend of the CQP web (Hardie, 2012) that hosts various corpora for the public and academics, and the commercial corpus tool, SketchEngine (Kilgarriff et al., 2014). CWB was used for this corpus, instead of the more common and user-friendly tools such as Antconc (Anthony, 2017) and Wordsmith (Scott, 2016) because the corpus in this thesis is relatively huge. Antconc and Wordsmith both require data to be loaded every time they open, and it is time-consuming and unstable for big data, whereas CWB only requires the data to be encoded once and can manage large corpora ranging from 10 million to 2 billion words.

More importantly, CWB allows a data structure that can authentically encode the online discussions. The comments in Futurelearn are nested within four levels: Courses, Steps, Threads, Comments, as visualized in Figure 5.3, and are encoded in XML according to TEI standard (TEI Consortium, 2020). All these levels are annotated and used in query for frequency counts, concordancing and collocation. This feature is essential for sub-corpora comparison of the keyword analysis in this thesis.

Figure 5.3 Nested structure of online discussions in FutureLearn



Note. This shows an example of a thread with id 4509705 consisted of an initiating post and two replies in step 1.08 in the dyslexia-1 course.

5.2.4 Statistical tool and qualitative analysis tool

CWB does not provide statistical analysis for keyword analysis as in other corpus tools. Therefore, R programming was used to conduct the statistical analysis. It was achieved by importing the output from CWB on word frequency and collocates frequency into R, then conducting statistical analysis following the method and criteria described in section 4.5. The descriptive analysis of the corpus in section 5.3 was also conducted in R.

The output of the CWB is not easy to work with for qualitative analysis, although it allows random sampling, sorting and widening co-text in concordancing to reveal the complete thread. QSR International's NVivo 11 and R programming were used instead for organizing threads for micro-analysis of threads. The keywords were queried in the CWB, the id of the comments or threads where the keywords occur were then retrieved. The ids were then used in R to extract the full threads or comments to be displayed in NVivo for qualitative analysis.

5.3 Ethical considerations

As this thesis analyses online discussions in FutureLearn, ethics must be considered following the guidelines of both FutureLearn and academic research. Before reviewing the guidelines, the nature of the analysis and the public-private domain distinction of online discussions are first discussed. Decisions taken in this thesis regarding informed consent and confidentiality are then justified in relations to these concerns and guidelines. This thesis received ethical approval from the Open University Human Research Ethics Committee (Appendix B).

The nature of data analysis

In this thesis, comments on FutureLearn, which constitute the data for this project, are provided in a format without any identifiable personal data, as shown earlier in Table 5.3. Both quantitative analysis and qualitative analysis were conducted in this thesis. The quantitative analysis aggregated all users' data and it was not possible to identify any individuals from the analysis. The qualitative analysis, especially the micro-analysis, required me to read into what had been posted, and present the threads in this thesis. Most of the time, I could conduct the micro-analysis with only the anonymized data. However, there were times I had to visit the platform to pinpoint which post or reply users referred to when they addressed another user. In this case, I did not work with anonymized data but saw the users' name beside their comments. In-depth reading of users' comment on sensitive issues, political stances, personal experience or health issues borders on revealing their privacy (European Commission, 2018; Eysenbach & Till, 2001).

Public-private domain distinction of online discussions

Online discussions in MOOC straddle the private and public domains. Unlike closed online communities or mailing lists, the online discussion in each FutureLearn MOOC is open and visible to all users as long as they hold an account with FutureLearn and register as a learner for that particular MOOC. This means that one's comment can be read by thousands of users, and is to some extent publicly available. For example, a comment in nutrition-4 course can potentially be read by 12109 users who have registered with the course. Despite this potentially large audience for one's

comment, however, the online discussion in FutureLearn is unlike online news websites or social media such as Twitter that allows anyone to read the comments or any crawler to discover its content. Only users taking the course are allowed to read the online discussions. Thus, the online discussion in FutureLearn can be private in nature.

Informed consent

The nature of the analysis conducted in this thesis and the blurred distinction of private-public domain need to be considered when deciding on the need for informed consent and confidentiality. Informed consent is not sought explicitly in this thesis for three reasons. Firstly, according to FutureLearn research ethics guidelines (FutureLearn, 2014), users are informed that their activities are monitored for research purposes when they sign up for the MOOC, so an opt-in consent is not needed, although users can opt out by unregistering from FutureLearn. Secondly, the potential harm towards the users is minimal. I did not intrude into the on-going discussions as the data was collected and analysed retrospectively (Eysenbach & Till, 2001). Also, the data collection did not take up users' extra effort because the data was generated spontaneously (Jaworska, 2018). Furthermore, it is expected that the publication of the results will not cause harm to any of them if confidentiality is maintained (to be explained next). Thirdly, it is not pragmatic to contact every user due to their large number and some of the MOOCs ended before this project started. Furthermore, if I were to approach them for consent, I would have to associate the anonymized user id with their real name, which is not allowed according to FutureLearn privacy policy (FutureLearn, 2019). Admittedly, although users are informed by FutureLearn regarding the possibility of research, they might want to be informed of the nature of a particular study and be given the chance to opt-out. Also, according to General Data Protection Regulation (GDPR) which was introduced after the data collection ended, users must actively opt in instead of by default. However, given the abovementioned reasons, I decided not to seek informed consent, while ensuring users' confidentiality to minimize any harm or unsatisfaction.

Confidentiality

Confidentiality is to protect the privacy of users. It is achieved by analysing anonymized data, and storing all the data in password-protected systems. However, confidentiality becomes an issue when the exact threads or comments are presented for the qualitative analysis of this thesis.

According to FutureLearn terms and conditions (FutureLearn, 2014), users' comments are treated as an intellectual property and subjected to a Creative Commons Licence (Attribution-Non Commercial-NoDerivs; BY-NC-ND). Under this scenario, users should be acknowledged for their comments if they are being quoted. However, this conflicts with confidentiality. This conflict arises because researchers, especially linguists, treat users' comments or discourse as data, rather than an idea that can be publicized, sold, copied or referenced. Therefore, most linguistic research on public online discussions prioritize confidentiality over acknowledgement (e.g., Baker & Egbert, 2018; Jaworska, 2018). Acknowledgment with user's name may compromise anonymity and potentially pose a threat to the user if the comments quoted involve sensitive issues such as health, personal experience or political stance (British Psychological Society, 2017).

Additionally, quoting with user's name in settings other than where it originally occurs, i.e., a thesis or research publications, assumes that their contributions in the online discussions of FutureLearn are publicly available and they are aware of it. However, as discussed earlier, the comments within the FutureLearn are only visible to those registering with the MOOC. It is possible that the contributing users may only intend their comments to be read by fellow users in the course, rather than by researchers or readers of a research publication. This is in contrast to other online users such as bloggers, YouTubers, and activists who intend their content to be spread among the public (British Psychological Society, 2017).

In this thesis, a solution is devised to acknowledge users while maintaining their confidentiality. Whenever a thread or comment is presented in the findings, instead of the name of the contributing user, acknowledgment is achieved by showing the URL linked to the comment, for example <https://www.futurelearn.com/courses/moons/1/comments/769923>, in the footnotes. This

URL links to the FutureLearn platform where the comment can be found, along with the contributing user's name. However, only users registering with the same presentation of the course can view the page. Users' confidentiality can thus be maintained while being acknowledged. This is also in line with the assumption that online discussion in FutureLearn is private in the sense that users only contribute posts for fellow users to read and should only be identifiable to them. Admittedly, their wording is still presented in this thesis other than the original setting, but it is believed that the harm towards them is minimal given that they are not identifiable to people other than fellow users in the same course. As mentioned earlier, users' comments are used as data in this thesis, rather than sources or references, so they are not acknowledged according to academic referencing and citation. Lastly, given that users are anonymized when their comments are presented in the findings, the gender-neutral pronoun *they* is used to refer to a specific user instead of gender-specific *he* or *she*.

5.4 The FL Corpus: Further descriptions

At the start of this chapter, the comments and wordcount of the sub-corpora of the FL corpus to be subjected to corpus analysis have been introduced. In this section, I further describe the corpus on several aspects including the MOOCs, length of threads, and users' contributing patterns to provide a background to the online discourse to be examined.

5.4.1 MOOCs in the corpus

The details of each MOOC included in the corpus are summarized in Table 5.4. The 12 MOOCs are of different disciplines, including history, sciences, and social sciences. The MOOCs also vary in length in terms of the weeks and number of the steps on which users can comment in the online discussions. Admittedly, these 12 MOOCs are in no way representative of the variety of MOOCs offered on FutureLearn. However, with a wordcount of 11 million, the corpus is considered relatively large, comparable to the 11.5-million-word spoken component of BNC2014 (Love et al., 2017).

Therefore, it can be safely assumed that it would capture general patterns and possibly rare but important discourse practices in the MOOC online discussion.

Table 5.4 Summary of the 12 MOOCs included in the corpus

MOOC	Abbreviation	Presentation	Date when the first comment is posted	Offering University	Course Slogan	Number of Course Week	Number of Steps that have discussion space
Digital Accessibility: Enabling Participation in the Information Society	accessibility-2	2	2/6/2017	University of Southampton	With a better understanding of users' needs, technologies can be developed to be accessible & provide a more inclusive environment	5	84
Health and Wellbeing in the Ancient World	ancient-1	1	2/6/2017	The Open University	Discover what healthcare was like in ancient Greece and the Roman world with this free online course.	6	94
Learn to Code for Data Analysis	code-1	1	10/26/2015	The Open University	Software and data make the world go round. Learn programming, to analyse and visualise open data, with this free online course.	4	98
Corpus linguistics: method, analysis, interpretation	corpus-1	1	1/27/2014	Lancaster University	Offers practical introduction to the methodology of corpus linguistics for researchers in social sciences and humanities	8	267
Dyslexia and Foreign Language Teaching	dyslexia-1	1	4/20/2015	Lancaster University	Gain practical tools and theoretical insights to help dyslexic students learn second languages, with this free online course.	4	51
Inequalities in Personal Finance: the Baby Boom Legacy	finance-1	1	3/23/2015	The Open University	Explore the concerns about rising generational and economic inequalities in developed countries, with this free online course.	4	84

Contract Management: Building Relationships in Business	management-4	4	11/14/2016	University of Southampton	Learn to build relationships and manage contracts successfully with this free online course backed by UK government and IACCM.	3	59
Moons	moons-1	1	3/17/2014	The Open University	Explore the many moons of our Solar System. Find out what makes them special. Should we send humans to our Moon again?	8	206
Nutrition and Wellbeing	nutrition-4	4	1/16/2017	University of Aberdeen	Demystify the complex and conflicting messages we hear about nutrition, health and lifestyle today, on this free nutrition course.	4	52
Exploring our oceans	oceans-1	1	2/3/2014	University of Southampton	Explore the half of our world covered by deep ocean, and how our lives affect the hidden face of our planet.	6	81
Palliative Care: Making it Work	palliative-1	1	10/17/2016	Lancaster University	Learn how palliative care is managed in Europe and find out about best practice in delivering integrated palliative care	3	38
Soils: Introducing the World Beneath Our Feet	soils-1	1	7/6/2015	Lancaster University	Learn about soils, the variety of life they contain and how humans impact this fragile system, with this free online course.	4	42

5.4.2 Users' comments and wordcount in each MOOC

In Futurelearn, only a proportion of users contribute to the online discussion while most users go through the MOOCs without contributing. The corpus only concerns those who contribute comments because the focus of this thesis is on online discourse. The percentage of contributing users out of all users taking the MOOCs can be found in Table 5.5.

Table 5.5 Users' comments and wordcount across the 12 MOOCs.

Abbreviation	Number of contributing users	Number of all users taking the MOOC ¹	Proportion of contributing users out of all users	Comments contributed by users	Average number of comments per user	Wordcount of users' comments	Average wordcount per users' comment
accessibility-2	609	2751	22%	7848	2.85	444440	56.63
ancient-1	994	3751	26%	23108	6.16	1357658	58.75
code-1	1968	8468	23%	10310	1.22	418999	40.64
corpus-1	2198	5470	40%	14518	2.65	1084845	74.72
dyslexia-1	5722	9955	57%	41175	4.14	2325367	56.48
finance-1	637	1952	33%	10033	5.14	820829	81.81
management-4	934	4337	22%	8415	1.94	444294	52.80
moons-1	1985	5115	39%	24038	4.70	927806	38.60
nutrition-4	3788	12109	31%	30923	2.55	1443443	46.68
oceans-1	1336	5237	26%	7557	1.44	374628	49.57
palliative-1	1320	3012	44%	10518	3.49	706377	67.16
soils-1	2004	4345	46%	14344	3.30	792637	55.26

¹ This refers to users visiting at least one step of the MOOC

As shown in Table 5.5, the wordcount and number of users' comments varies across the MOOCs. This is probably because, as we saw in Table 5.4 above, the MOOCs are of different length and involve different numbers of contributing users. The corpus may be imbalanced because some MOOCs contain double the wordcount of others, such that the corpus may reveal mainly the pattern of those courses (McEnery & Hardie, 2012). This imbalance is taken into account with the dispersion measure in the keyword analysis, as described in section 4.5.3.

5.4.3 Sub-corpora of users' comments for keyword analysis

The corpus information regarding the sub-corpora of users' initiating posts, independent posts and replies that are to be compared in the keyword analysis is presented in

Table 5.6. Overall, there are 3.7 times more independent posts than initiating posts. In all but corpus-1, there are more independent posts than initiating posts. There are more initiating posts than independent posts in corpus-1 probably because 22 facilitators actively reply to the users in this course. Although the number of replies is 1.65 times more than initiating posts, the number of replies is 5.33 times fewer than initiating posts and independent posts collectively, suggesting users tend to create new post rather than replying to others.

Table 5.6 Number of comments and wordcount of the three types of comments in the corpus.

	<u>Initiating post</u>		<u>Independent post</u>		<u>Reply</u>	
	Number of Comments	Wordcount	Number of Comments	Wordcount	Number of Comments	Wordcount
accessibility-2	1151	92680	4731	262140	1966	89620
ancient-1	4176	330828	8652	567627	10280	459203
code-1	1900	105238	4822	169263	3588	144498
corpus-1	5496	500678	4443	297091	4579	287076
dyslexia-1	4841	406832	30792	1652576	5542	265959
finance-1	1845	188831	2612	232226	5576	399772
management-4	1207	73449	5905	310497	1303	60348
moons-1	4386	201854	11724	420635	7928	305317
nutrition-4	2661	197137	23109	1024804	5153	221502
oceans-1	1626	92151	4278	214175	1653	68302
palliative-1	720	70113	8525	572934	1273	63330
soils-1	2043	140227	8211	435620	4090	216790
Total	32052	2400018	117804	6159588	52931	2581717

In this corpus, 75% of the new posts do not receive replies. The proportion of independent posts are more than what have been found in previous studies on other online spaces (Beth et al.,

2015; Meyer et al., 2019; Napoles et al., 2017). This suggests that users tend to be *prompt-focused* (Herring, 2013), such that they respond to the prompt in the course by creating posts, rather than replying to others' comments. Although large number of independent posts indicate users' engagement in the MOOCs, the vast difference in the number of the initiating posts and independent posts also further attest to the importance of examining the discourse practices characterizing these two types of posts, in order to determine what kinds of discourse practices invite replies, thus initiating user-user interactions. The keyword analysis of initiating posts and independent posts will be conducted in Chapter 6 and 7.

A reply can be further differentiated based on whether it is contributed by a user who just joins the thread, i.e., their first contribution, or by a user who has been involved in the thread either by posting an initiating post or a reply before, i.e., their subsequent contribution. The latter provides an approximate measure of users' recurrent interaction and continued engagement in the same thread, where negotiation is likely to occur. The distributions of these two types of replies is presented in Table 5.7, and it shows that subsequent contributions are 2.37 times fewer than first contributions, suggesting users come back to the thread that they contribute before but not often. Users' continued engagement within the same thread will be further explored in the micro-analysis of threads in Chapter 8, especially when users hold opposing stances.

Table 5.7 Number of replies and wordcount of two types of replies in the corpus.

	<u>First Contributions</u>		<u>Subsequent Contributions</u>	
	Number of Replies	Wordcount	Number of Replies	Wordcount
accessibility-2	1472	64342	494	25278
ancient-1	6980	296042	3300	163161
code-1	2533	100466	1055	44032
corpus-1	2177	129212	2402	157864
dyslexia-1	3858	178022	1684	87937
finance-1	3897	269475	1679	130297
management-4	1015	47359	288	12989
moons-1	6575	248939	1353	56378
nutrition-4	3816	151216	1337	70286
oceans-1	1185	49800	468	18502
palliative-1	1011	49500	262	13830
soils-1	2717	129656	1373	87134
Total	37236	1714029	15695	867688

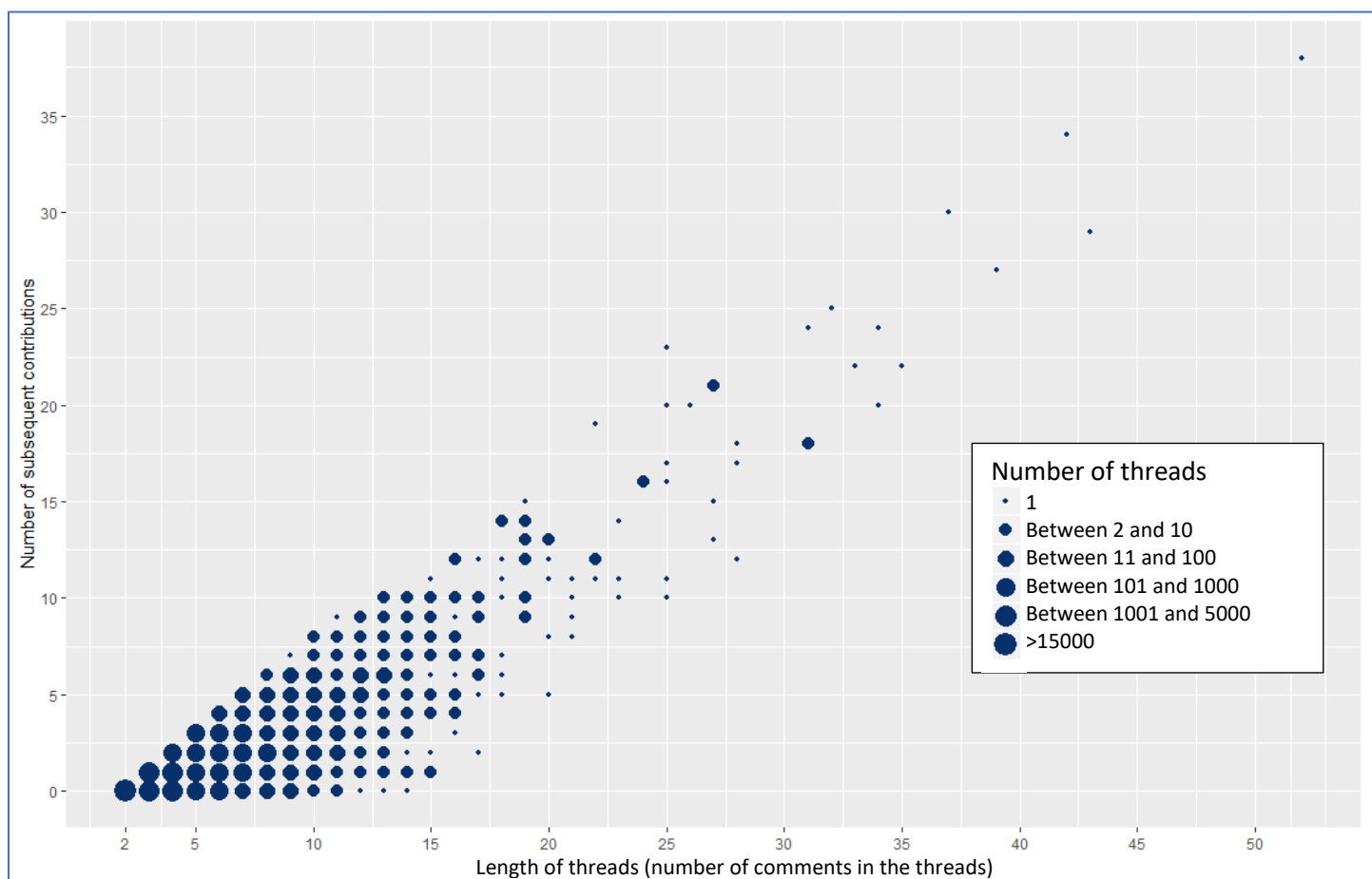
5.4.4 Length of threads

Threads, i.e., the initiating post and replies it receives, are where explicit user-user interactions occur (R. Ferguson & Sharples, 2014; Ksiazek & Lessard, 2016; Lewis, 2005). The minimum length of a thread is two. The number of threads of different length across the 12 MOOCs is presented in Figure 5.4. The total number of threads are 32334, 32052 of which are initiated by users and 281 initiated by facilitators. The length of the threads in the corpus is highly skewed, as fifty-one percent of the threads are with only one reply, i.e., with a length of two. This finding corresponds to other MOOC platforms (Cui et al., 2017) and YouTube commenting spaces (Bou-Franch, 2012). The overabundance of one-reply threads, as well as the independent posts, speaks to the possibility that users tend to respond to prompts rather than engage in sustained discussions (Herring, 1999, 2013). Beside one-reply threads, another 40% of the threads consist of 2 to 4 replies. I categorize all these threads, i.e., threads consisted of 1 to 4 replies, as *short threads*, and together they account for 91% of the threads in the corpus.

The remaining 9% of the threads range from 5 to 51 replies, and are categorized as *long threads*. Among these long threads, 86% comprise 6 to 10 comments, 12% 11 to 20 comments, and

The categorization of the *short vs long* threads is admittedly arbitrary, but Savolainen (2001) also found similar trends with the same categorization in newsgroup such that long threads are rare. It is possible that in long threads users are more likely to engage in continuous discussions, turn-taking and negotiation (Lapadat, 2002; Tubman et al., 2016). One piece of empirical evidence supporting this assumption can be drawn from the current corpus. As shown in Figure 5.5, there is a correlation, $r=.85$, between the length of the threads and the number of subsequent contributions in the threads. The high number of subsequent contributions in a thread suggest that users may be responding to each other, thus engaging in turn-taking and negotiation.

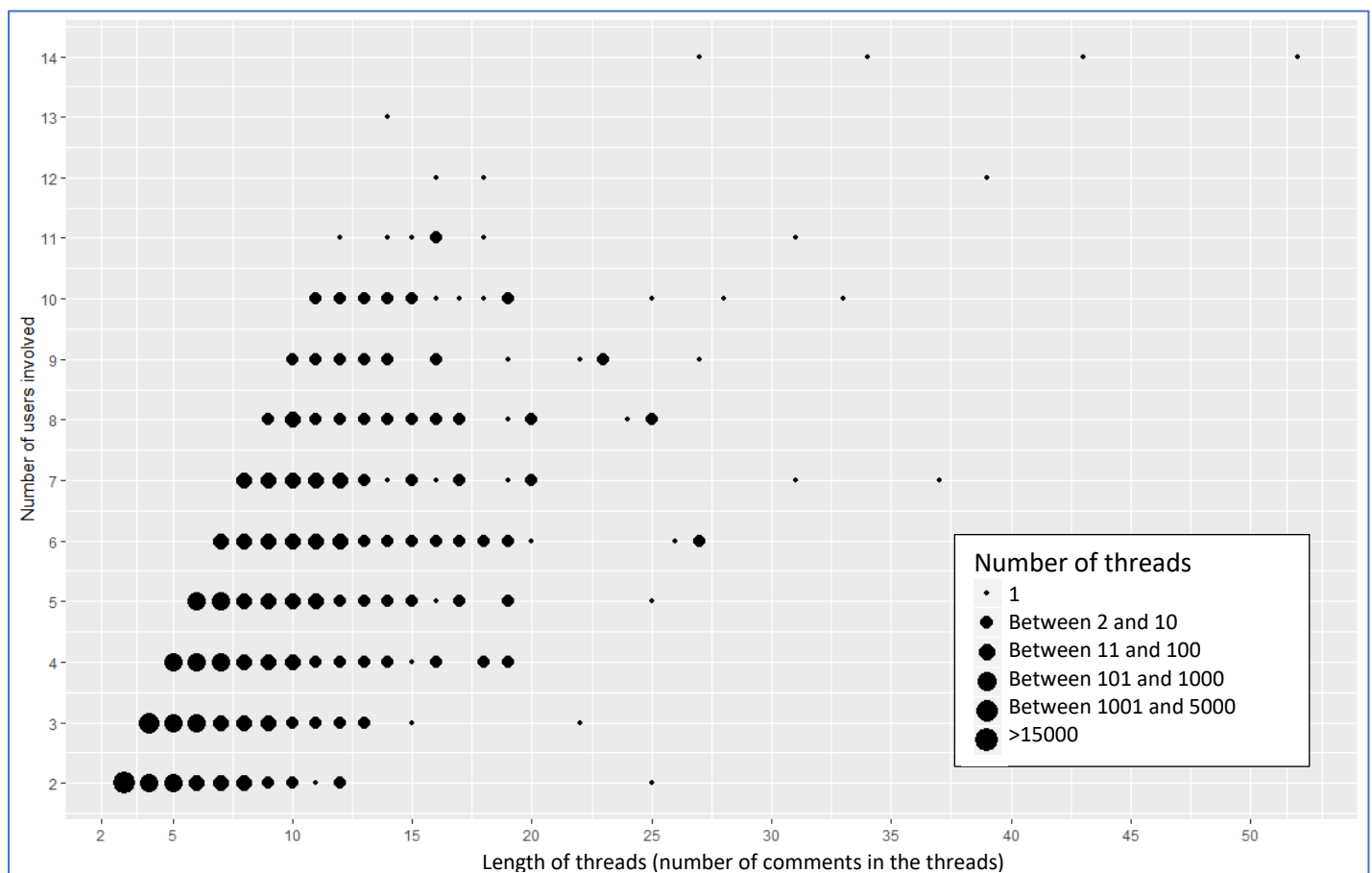
Figure 5.5 Scatterplot: Number of subsequent contributions vs. length of threads



Note. The longer the threads, the more subsequent contributions they contain. The size of the dots indicates the number of threads that correspond with a particular length and number of subsequent replies.

However, there is also an exception. As shown in Figure 5.5, there are long threads comprising 5 to 14 comments without any subsequent contribution such that each user only replies once in the thread. One such thread of 9 replies is shown in Chapter 8, where the users mainly reply with agreements. This observation possibly is associated with users who just post and go, and do not come back to the threads initiated by themselves or the threads they have replied in before (see section 5.4.5 for users' contribution patterns). Furthermore, in only 24% of the threads, the initiator who starts the thread comes back to make subsequent contributions, suggesting that not every initiator engage in conversations they initiate. Besides subsequent contributions, the long threads are also more likely to involve many users, thus where polylogue occurs, as shown in Figure 5.6. Together, Figure 5.5 and Figure 5.6 point to the complex user-user interactions within a thread. However, this quantitative analysis does not reveal how the threads evolve. As argued in Chapter 3, this detail will only be revealed in micro-analysis of threads. For example, in the thread with 41 replies, two users each contribute 17 replies and mainly address each other, although there are 11 other users reply in the thread and the initiator never come back despite so many replies in the thread.

Figure 5.6 Scatterplot: Number of users involved vs. length of threads



Note. The longer the threads, the greater the number of users are involved. The size of the dots indicates the number of threads that correspond with a particular length and number of users involved.

5.4.5 Contributing users

As mentioned earlier, the background of the users was not collected in this thesis because these are archived data and very few users volunteered to provide this information in response to the survey at the beginning or end of the MOOCs. However, based on previous research on FutureLearn (Liyanagunawardena & Williams, 2016; Rizvi, Rienties, & Khoja, 2019; Shi & Cristea, 2018; Swinnerton et al., 2017), it can be assumed that most users are based in the UK, degree holders, and more likely to be elderly, work part-time or not at all.

Although no demographic data is available, I will describe the contributing users for their posting patterns. I will present the descriptive analysis in terms of their frequency of posting and contribution patterns of different types of comments. I argue that this behavioural information is

more important than a categorical grouping of users based on their nationality, ethnicity or education background in understanding user-user interactions at the discourse level because their posting behaviours are what is exhibited in the online spaces (Jones & Hafner, 2012). However, I am not dismissing the demographic background. Rather I will follow a CA approach such that this information will only be relevant if it is raised within users' discourse (Antaki & Ardévol, 2005). It is also worth noting that the research goal of this thesis is to examine language in different types of comments and how language is used to initiate, sustain and hinder their conversation with others, and the role of demographic factors lies beyond the current scope.

5.4.5.1 Frequency of posting: Super-posters and one-time posters

There are 23495 users contributing comments in the 12 MOOCs in the corpus. Figure 5.7 gives an overview of the number of contributions by each user in each MOOC. The measure is the number of comments per course step to account for the difference in length of course. As shown in the figure, there are only a few users in each course who contribute more than one comment per course step in each course, while 75% of users contribute fewer than 0.12 comments per course step, equivalent to one comment every ten steps. Those few highly prolific users can be considered as super-posters who, according to previous research, tend to dictate the discussion topic but can also exert positive impact in the online community (Huang et al., 2014; Lambiase, 2010; Poquet et al., 2018). The presence of super-posters amid users who post very few comments mirror other online spaces, such as Yahoo! Answers (Savolainen, 2014) and website forums (T. Graham & Wright, 2014).

Each dot represents one user

Topic	Comments per step (Number of dots)
accessibility-2	10
ancient-1	18
code-1	5
corpus-1	5
dyslexia-1	15
finance-1	15
management-4	10
moons-1	10
nutrition-4	15
oceans-1	5
palliative-1	10
soils-1	15

In this thesis, super-posters are defined as the ten most prolific users in each course, thus 120 of them in total. Admittedly, this is an arbitrary cut-off and, as can be seen from Figure 5.7, non-super-posters in some courses (e.g., ancient-1) are more prolific than super-posters in other courses. Based on this definition, 8% of users' comments and 10% of wordcount are contributed by these 120 super-posters. Table 5.8 shows the different types of comments that super-posters contribute.

Table 5.8 Super-posters contributions in each MOOCs

	Initiating post		Independent post		Reply			
					First contribution		Subsequent contribution	
	Number of comments	% ¹	Number of comments	% ¹	Number of comments	% ¹	Number of comments	% ¹
accessibility-2	249	22%	511	11%	580	39%	204	41%
ancient-1	265	6%	494	6%	1420	20%	974	30%
code-1	74	4%	209	4%	367	14%	132	13%
corpus-1	612	11%	357	8%	311	14%	428	18%
dyslexia-1	132	3%	395	1%	444	12%	168	10%
finance-1	222	12%	155	6%	1038	27%	609	36%
management-4	84	7%	490	8%	103	10%	19	7%
moons-1	445	10%	941	8%	905	14%	382	28%
nutrition-4	91	3%	420	2%	518	14%	145	11%
oceans-1	147	9%	285	7%	276	23%	122	26%
palliative-1	45	6%	318	4%	234	23%	74	28%
soils-1	92	5%	144	2%	558	21%	234	17%
Total	2458	8%	4719	4%	6754	18%	3491	22%

¹ the percentage indicate the proportion of super-posters' comments out of all users' comments

As shown in Table 5.8, super-posters seem to be responsive to others, as indicated by their disproportionate number of replies. Half of the MOOCs contain more than 20% of first contributions by super-posters. Furthermore, half of the MOOCs contain more than 25% of subsequent contributions from super-posters, suggesting that super-posters tend to engage in continued discussions with others. This finding suggests the potential role of super-posters in minimizing the lack of replies and sustaining a conversation, parallel findings on other MOOC platforms (Huang et al., 2014; Poquet et al., 2018) that found that super-posters can be helpful in maintaining activity levels in online discussions. Although not the main focus of this thesis, the discourse practice of one super-poster who are active in replying to others will be highlighted in the micro-analysis of threads.

One-time posters

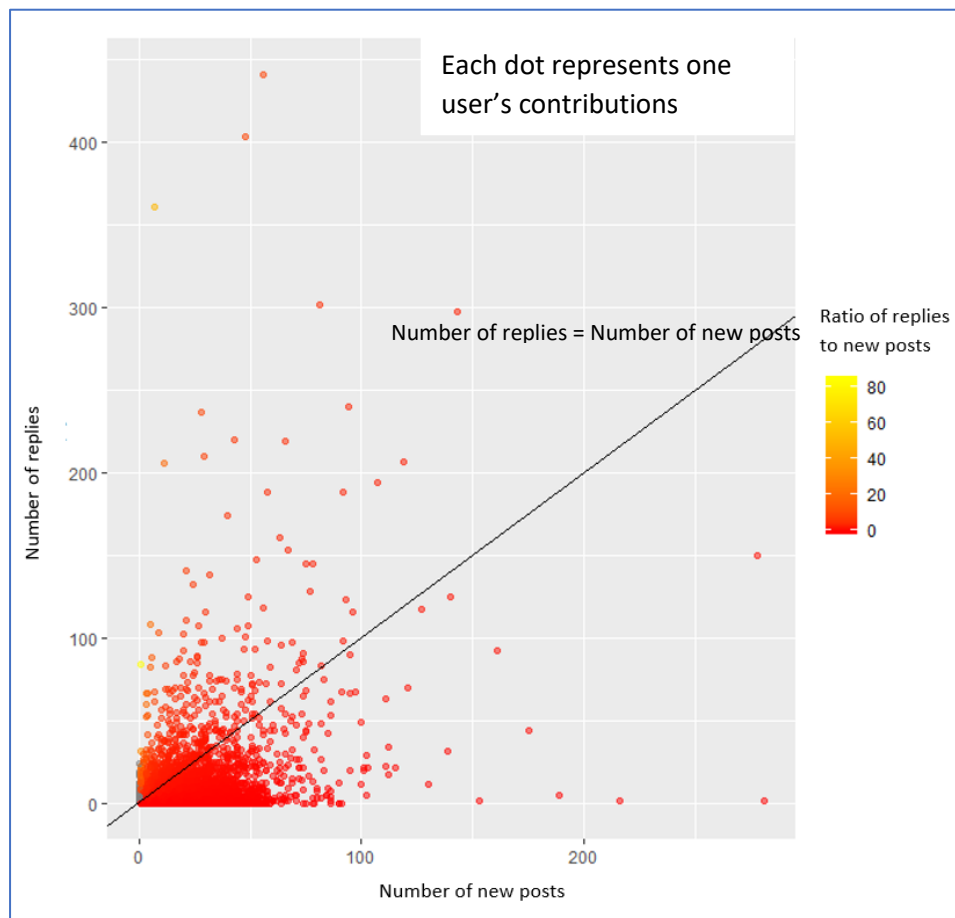
On the other end of the spectrum, 7960 users only contribute one comment, and another 7987 only contribute between 2 to 5 comments. The observation that users only posting once is also found in other online discussion spaces (Bou-Franch & Garcés-Conejos Blitvich, 2014; Herring, 1999;

Ruiz et al., 2011; Savolainen, 2011). It is possible that these users drop out of the course, which is common in MOOCs (R. Ferguson & Clow, 2015), or are put off for posting again because of not receiving replies from others for their first post (Herring, 1999; Joyce & Kraut, 2006). The latter hypothesis corresponds to the observation that 86% of them contribute a new post which does not receive any reply. These one-time posters thus do not establish any connections with others or engage in conversation with others.

5.4.5.2 Users' contribution patterns: Seven types of users

The users' contribution patterns in the online discussions can be identified by the extent they create new posts, whether initiating posts or independent posts, and reply to others. As mentioned earlier, replies to others can be further differentiated into their first contributions in a thread initiated by others, or their subsequent contributions to threads initiated by others or themselves. Figure 5.8 first shows the ratio of each user's contributions of new posts and replies, then Figure 5.9 illustrates detailed differentiation of users based on different types of replies and posts contributed.

Figure 5.8 Each user's contributions of new posts and replies



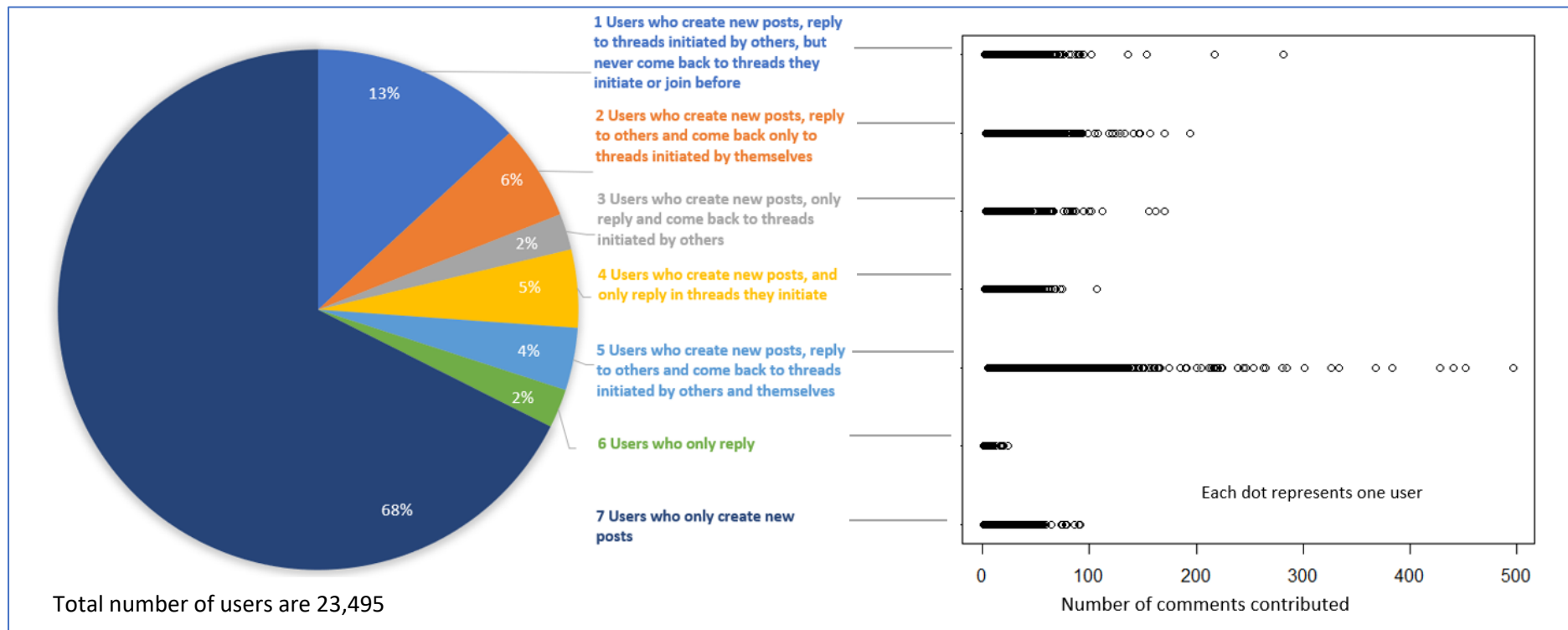
Note. The reference line is when a user contributes equal number of replies and new posts.

As shown in the upper left-hand corner of Figure 5.8, some users tend to reply more often than to create a new post. Nonetheless, only 1776 users contribute more replies than posts, whereas 791 users contribute equal number of replies and posts, and 20928 contribute more new posts than replies, of which 7576 of them only create new post once and contribute no reply at all. Those who contribute more replies than new posts may help to minimize the problem of lack of replies. As argued earlier, users tending to create posts rather than replying to others can be a *prompt-focused* behaviour in response to the content in the course step (Herring, 2013). Qualitative analysis of the independent keywords in Chapter 7 will partly reveal what lies behind this

observation – they engage in user-content interactions instead of user-user interactions in the online discussion of FutureLearn.

Figure 5.9 illustrates users' contributions in the online discussion based on whether they create new posts (whether initiating or independent posts), reply to others (first contributions), and make subsequent contributions in a thread they have initiated or replied to before. On one hand, making subsequent contributions in the same thread may indicate turn-taking given that users come back to the thread, thus possibly a sustained conversation. On the other hand, creating new posts and replying but never making subsequent contributions may indicate commenting and leaving, rather than having a sustained conversation.

Figure 5.9 Different types of users categorized based on their contributions



Note. The pie chart shows the proportion of different types of users whilst the plot shows the number of comments made.

Seven categories of users are identified. Firstly, 68% (n=15875) of users only create new posts without replying to others (Category 7 in the figure⁸), although 7576 of them only post once. 88% of their posts are independent posts that do not receive replies. Interestingly, among these users who only create new posts, except those who only post once, 5279 of them continue posting (number of new posts contributed ranges from 2 to 55), despite not receiving any reply for any of their posts, suggesting they might be posting to engage with the content, rather than with others. This might also explain why they do not reply to others. Secondly, two percent of users only reply to others but never create a post (Category 6) and make very few contributions to the discussions.

The other five groups of users all create posts, whether initiating or independent posts, but differ in their replying activities. This difference is crucial for conversations among users because replies are explicit action that users take to engage in user-user interactions (Ksiazek & Lessard, 2016; Lewis, 2005). One group of users (Category 5) reply to threads initiated by others and engage in continued interactions in the threads initiated by others or themselves. This group of users engage in all ways to participate in the discussions, probably resulting in several of them contributing more than 100 times in the discussions. Their active replying also suggests that they engage in conversations with others. In contrast, another group of users (Category 4) only reply in threads they initiate, and never reply to threads initiated by others. It can be speculated that they might not read others' post but only read others' reply to their threads. To some extent, they could be passive in establishing conversations with others because it is other users replying to them. Another group of users (Category 3) reply to others and engage in continued discussion in threads initiated by others, but never reply subsequently to the thread initiated by themselves. It is possibly by chance that they do not engage in continued discussion in threads initiated by themselves, given that they do engage continuously in threads initiated by others. This might be similar to the case of another group of users (Category 2) who reply to others' threads but only make subsequent contributions in threads

⁸ Due to the technical constraint of the graphing tool, I could not match the numbering in the figure to the order of my description in text.

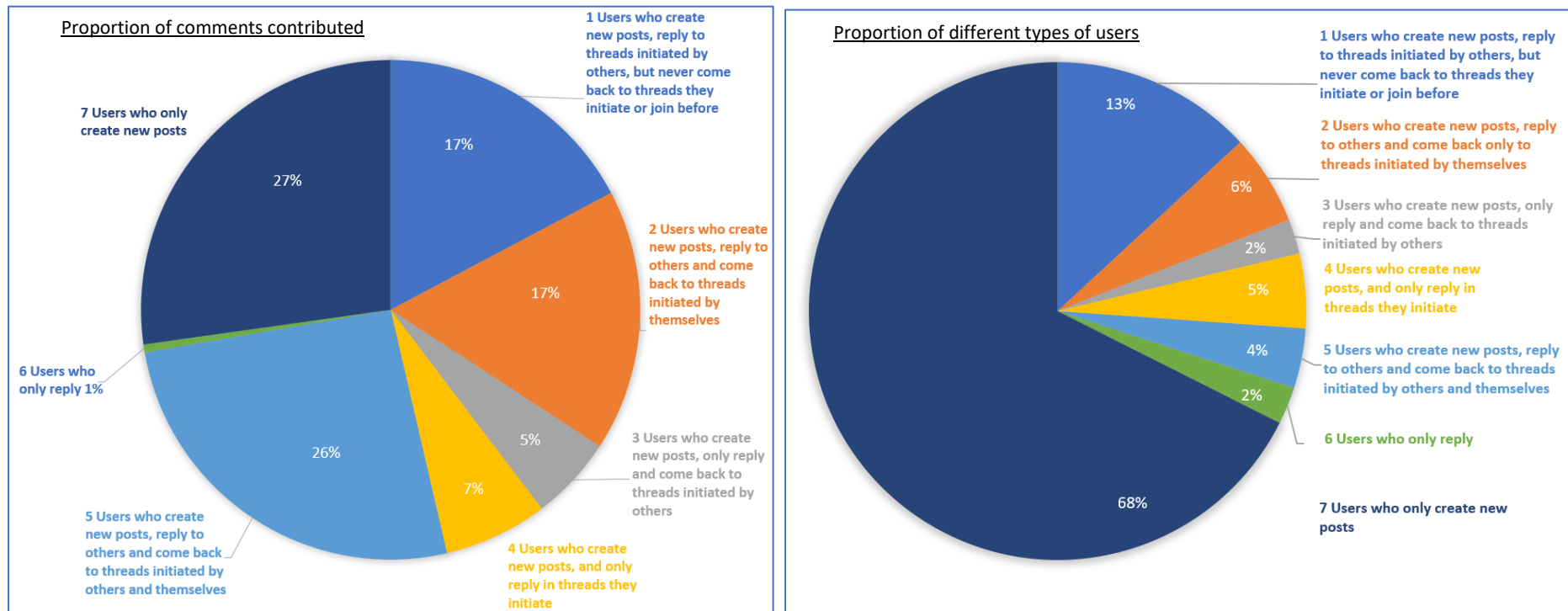
initiated by themselves. Lastly, there is a large group of users who never engage in continuous discussions in threads, whether initiated by others or themselves (Category 1). This could be similar to the prompt-focused posting such that they reply to others' posts and create new posts but do not come back to any threads they join before.

It could be said that users in Category 2 to 5 may have involved in some turn-taking with others in the threads given that they contribute more than once in the same threads. Also, it is worth noting that among the users in Category 1 and 3 who create new posts and reply to others, there are 1371 users whose new posts (new posts contributed ranges from 1 to 60 posts) never receive any replies, these are the independent posts that never start a thread that they could come back to. It is a shame that these users reply to others but never receive replies from others for their posts, and they might be different from those in Category 7 who only create new posts but never reply to others. They may be similar to those users who find user-user interactions in online discussions are compromised (Delahunty, 2018; Hew & Cheung, 2014; Hewings et al., 2009; Joyce & Kraut, 2006; Springer et al., 2015).

The proportion of comments contributed by each category of users as a group is presented in the pie chart in Figure 5.10, with the pie chart in Figure 5.9 which shows the proportion of each category of users reproduced alongside it. The proportion of comments contributed does not parallel the proportion of users. Although the largest group of users, i.e., the users who only create new posts (Category 7), still contribute the greatest number of comments, their proportion is only about one fourth of all the users' comments, probably because most of them contribute only one post. In contrast, users in the other categories, despite being fewer in number, also reply to others while creating new posts, thus contributing more comments collectively (Category 1 to 5). Among these groups, one group also contributes about one fourth of all users' comments. They are those who create new posts and reply to others, while also make subsequent contributions to the threads initiated by others and themselves (Category 5). Their different types of contributions make up a large proportion of all the comments. This is followed by those who create new posts and reply to

others but never make subsequent replies in threads they first reply to or initiate (Category 1) and those who only return to threads they initiate (Category 2). Lastly, the users who only reply contribute the least, probably because of their small number (Category 6).

Figure 5.10 Proportion of comments contributed by different types of users compared to the proportion of each type of users.



This analysis of users' contributions shows different commenting behaviors displayed by users in the online discussions. The categorization of users based on the differentiation between new posts and replies, subsequent contributions and first contributions in the threads seems to be valid rather than an arbitrary classification. Specifically, it shows that some users engage in turn-taking in a thread as if it is an oral conversation, whereas some users post, reply and leave the threads, similar to prompt-focused posting (Herring, 2013). Admittedly, it remains unknown whether users intentionally create posts only, reply only, only attend to the threads initiated by themselves, do not engage in continuous discussions or contribute all types of comments. It also remains unclear if users read others' responses to them even though they do not reply or make subsequent contributions in the threads they initiate or reply before. The qualitative analysis of reply keywords and micro-analysis of threads in Chapter 8 will explicate how some of these contributing patterns pan out. For example, there are times users come back just to say they do not want to engage in a discussion, two users continue their conversation within a long thread, and users whose new posts initiate long threads but never come back to make subsequent contributions despite others' replies.

5.5 Conclusion

This chapter illustrates how the 11-million-word FL corpus was created, from data collection to encoding to the corpus tool to corpus analysis. Specifically, I show that a large-scale corpus can be created, stored, annotated and analysed with the use of CWB tools, coupled with R and Nvivo, rather than being limited to the more common tools such as Antconc (Anthony, 2017) and Wordsmith (Scott, 2016) that are not efficient in processing big data.

The ethical consideration in using users' textual contributions highlights the tension between the public and private domain in online spaces, as well as between the view of hosts of the online spaces, in this case FutureLearn, and academic researchers regarding copyright and confidentiality. By taking into account these different views, a middle ground is reached by

presenting users' textual contributions with URLs linked to their comments, thereby the users' copyright is acknowledged following FutureLearn's requirement while their confidentiality is safeguarded following academic research guidelines. This ethical consideration is particularly relevant nowadays when academic researchers and technology companies collaborate in researching users' textual contributions and behaviours online.

This chapter also describes the FL corpus in detail. Despite a lack of demographic data, which is quite common in most online spaces where users are anonymous, the identification of users based on their contribution of different types of comments reveals that users exhibit different posting behaviours online. Firstly, there are two extreme groups, one-time posters and super-posters, as observed previously in various online spaces (Bou-Franch & Garcés-Conejos Blitvich, 2014; Herring, 1999; Huang et al., 2014; Lambiase, 2010; Poquet et al., 2018; Ruiz et al., 2011; Savolainen, 2011). The former is a large group of users who contribute once, and the latter are the minority group in the online discussions with massive contributions. On one hand, one-time posters will not be under-represented in the corpus given the large number of them. On the other hand, super-posters' massive contributions justify the possible over-representation of their textual contributions in the corpus analysis.

Secondly, besides frequency of posting as used by previous research, users posting behaviours can be further identified based on their replying behaviours, which to the best of my knowledge, have not been explored before. Although there are users who not only create posts, but also reply to others and engage in continuous discussions with others within a thread, most users simply create new posts without replying to others, or contribute more new posts than replies. This might partly explain why there are more new posts, especially independent posts, than replies in the corpus. This finding also suggests that users tend to be prompt-focused and create new posts in response to the course content. This finding will be further evidenced with keyword analysis of independent posts in Chapter 7 which draws attention to the user-content interaction encouraged by the design of FutureLearn. However, as argued in previous chapters, this observation may

disappoint those users who join online discussions for socialization and deliberation. To inform users who are keen to have a conversation with others, keywords analysis of initiating posts are conducted in Chapter 6 to understand how user-user interactions are initiated.

Thirdly, although there are many users who never come back to a thread they have initiated or replied to before, it is found that most long threads consist of subsequent contributions. This suggests that users respond to each other within a long thread, engaging in turn-taking, although the turns can be interleaved by other turns. A long thread might therefore be where negotiation and intersubjectivity happen. The existence of long threads provides a rare opportunity for understanding explicit user-user interactions through micro-analysis of threads, given that the majority of threads only have one reply, or fewer than five replies in the corpus, which is a common observation in online discussions (Cui et al., 2017; Napoles et al., 2017). The micro-analysis of threads will be conducted in Chapter 8 and 9, and will reveal disagreements as one of the occasions where users make subsequent contributions within the same thread and engage in intersubjectivity, rather than just posting and leaving the thread.

Having explored the users' posting behaviour in the FL corpus, it is time to investigate their textual contributions to understand the dialogic nature of online discourse in Chapters 6 to 9. Although the FL corpus compiled in this thesis consists of both users and facilitators' comments, only users' comments are analysed. The sub-corpora for corpus analysis are initiating posts, independent posts and replies, each of which will be the main focus for the next three chapters respectively.

Chapter 6

A Keyword Analysis of Initiating Posts vs. Independent Posts: Potential start of dialogic conversations

6.1 Introduction

This chapter and the next two chapters examine the textual evidence of the dialogic nature of online discourse in the three sub-corpora of the FL corpus respectively: initiating posts, independent posts, and replies. All investigations start from the keyword analysis of corpus linguistic approach.

This chapter, as well as the next chapter, explores the dialogic nature of online discourse from the point where a conversation can potentially start, that is the new posts. The keyword analysis comparing the two types of new posts, that is initiating posts that receive replies and independent posts that do not. This is based on the assumption that discourse practices in initiating posts start user-user interactions, thus potentially dialogic conversations, whereas discourse practices in independent posts could serve, in some cases, to realize user-content interactions, rather than establish conversations (Ksiazek & Lessard, 2016; Ziegele et al., 2014).

Following the keyword analysis, the functional grouping of initiating and independent keywords is then presented, thus beginning to address the first research question:

RQ1: What are the differences in the linguistic features and discourse practices that regularly occur in

- **initiating posts that receive replies and start a discussion thread,**
- **independent posts that do not receive replies.**

Then, a detailed description of selected groups of initiating keywords is presented to further explain the linguistic features and discourse practices that increase the chance of triggering conversations. Those in the independent posts are briefly mentioned in this chapter in comparison to initiating posts, and selected independent keywords will be explored in the next chapter. This chapter concludes by discussing the discourse practices for starting user-user interactions realized by the initiating keywords, thereby revealing online discourse that are of a dialogic nature. The discussion will thus start to address the second research question:

RQ2: How do these discourse practices initiate, sustain or hinder dialogic conversations in online discussions?

6.2 Keyword Analysis: Initiating Posts vs. Independent Posts

The keyword analysis is conducted by comparing the word frequency of every single word in the initiating posts and independent posts following the statistical procedure and criteria explained in section 4.5.3. Words that are used significantly more often in initiating posts than in independent posts are *initiating keywords*, and vice versa for *independent keywords*.

6.2.1 Initiating keywords

Sixty-nine keywords are found used significantly more often in the initiating posts than in the independent posts. The keywords are listed in Table 6.1 (see Appendix C for the full statistics of log-likelihood ratio test and dispersion measure). Given that the aim is to understand the difference between these two types of posts, the keywords are ranked by relative risk, an effect size measure that indicates how many times higher the normalized frequency of a keyword in initiating posts is, as opposed to that of in the independent posts.

Table 6.1 Initiating keywords ordered by effect size.

Keywords	Normalized Frequency in Initiating Posts ¹	Normalized Frequency in Independent Posts	Effect size ²	Keywords	Normalized Frequency in Initiating Posts	Normalized Frequency in Independent Posts	Effect size
please	25	6	4.42	perhaps	44	32	1.37
anybody	6	2	3.79	whether	37	27	1.36
wondering	14	4	3.16	might	75	56	1.35
?	599	208	2.88	used	134	101	1.33
anyone	36	13	2.68	rather	52	39	1.33
question	53	24	2.21	-	306	231	1.32
missing	11	5	2.15	example	55	42	1.32
"	629	293	2.15	ca	47	36	1.31
wonder	41	21	1.98	any	134	103	1.29
'	597	308	1.94	two	63	49	1.29
explain	15	8	1.89	he	96	75	1.29
surely	19	10	1.88	;	192	150	1.29
sorry	16	9	1.84	say	61	48	1.27
numbers	12	7	1.79	here	79	63	1.25
why	83	47	1.74	seems	66	53	1.25
:	242	141	1.72	if	308	248	1.24
says	16	10	1.69	then	128	103	1.24
does	118	70	1.68	did	124	101	1.23
told	18	11	1.68	same	88	72	1.22
mean	31	19	1.68	...	182	149	1.22
tell	20	12	1.66	than	159	131	1.22
sort	16	10	1.61	one	239	198	1.20
article	38	24	1.58	by	278	233	1.20
e.g.	22	14	1.57	there	323	275	1.18
came	21	13	1.57	n't	346	297	1.17
(477	308	1.55	just	163	141	1.16
called	22	14	1.53	was	505	436	1.16
else	24	16	1.52	could	185	160	1.16
wrong	25	16	1.51	's	345	303	1.14
)	543	360	1.51	would	382	337	1.13
1	55	37	1.51	or	429	379	1.13
tried	29	20	1.48	the	4375	3964	1.10
cannot	30	21	1.46	on	603	560	1.08
were	186	130	1.43	that	1105	1033	1.07
				,	3055	2908	1.05

Note. Only normalized frequency and effect size are presented here to reveal the differences between the two types of posts. The full statistics for the log-likelihood ratio test and dispersion measure (i.e., the criteria to decide whether a word is a keyword) of each initiating keyword can be found in Appendix C.

¹The normalized frequency is measured by per 100,000 words and is rounded so there is no decimal places.

²The effect size is measured by relative risk, that is the ratio of the normalized frequency in initiating posts to the normalized frequency in independent posts. The calculation is based on unrounded normalized frequency rather than the rounded one shown here.

As shown in Table 6.1, the normalized frequency of the initiating keywords varies, ranging from 6 occurrences per 100,000 words for *anybody* to 4375 occurrences per 100,000 words for *the*. Although the word frequency of some keywords is low, they are still significant given the fact that they are used more often in the initiating posts than the independent posts. The top eight keywords have an effect size above 2, i.e., they are used twice as often in the initiating posts than in the independent posts. Seven of these eight keywords, *please, anybody, wondering, ?, anyone, question, missing*, are typically used in seeking information, and will be further explored in section 6.3. At the bottom of the table are keywords with smallest effect size, such as *was, could, 's, would, or, the, on, that*, and the comma (,). The effect size is small for they are used frequently in both initiating posts and independent posts (range from 160 to 4375 per 100,000 words), probably because they are grammatical words. Despite their small effect size, modals such as *could* and *would* have been well-established as crucial in language communication and discourse analysis (Fairclough, 2003; Stubbs, 1986), and are thus explored in section 6.3.

6.2.2 Independent keywords

Seventy-seven keywords were used significantly more often in the independent posts than the initiating posts. The independent keywords are listed in Table 6.2

Table 6.2 Independent keywords ordered by effect size.

Keywords	Normalized Frequency in Independent Posts ¹	Normalized Frequency in Initiating Posts	Effect Size ²	Keywords	Normalized Frequency in Independent Posts	Normalized Frequency in Initiating Posts	Effect Size
joined	8	3	2.67	week	102	72	1.42
informative	22	9	2.44	agree	38	27	1.41
forward	71	31	2.29	feel	85	61	1.39
improve	31	14	2.21	easy	39	28	1.39
keen	11	5	2.20	very	361	263	1.37
hoping	14	7	2.00	lot	118	86	1.37
everyone	71	36	1.97	'm	221	162	1.36
knowledge	67	35	1.91	every	49	36	1.36
meet	17	9	1.89	love	46	34	1.35
currently	28	15	1.87	my	661	497	1.33
achieve	13	7	1.86	!	496	375	1.32
affects	9	5	1.80	working	66	50	1.32
opportunity	21	12	1.75	really	173	132	1.31
enjoyed	31	18	1.72	interesting	171	131	1.31
thank	84	49	1.71	better	78	60	1.30
definitely	29	17	1.71	difficult	67	53	1.26
gain	15	9	1.67	well	135	107	1.26
important	91	55	1.65	able	70	56	1.25
feeling	18	11	1.64	will	257	207	1.24
understanding	44	27	1.63	always	85	69	1.23
environment	37	23	1.61	our	181	147	1.23
enjoy	19	12	1.58	new	102	83	1.23
helps	22	14	1.57	think	271	221	1.23
looking	89	57	1.56	need	137	113	1.21
hope	53	34	1.56	and	2715	2264	1.20
yes	31	20	1.55	work	158	132	1.20
great	110	71	1.55	i	2905	2450	1.19
aware	32	21	1.52	more	404	342	1.18
good	187	123	1.52	about	342	290	1.18
education	38	25	1.52	also	257	219	1.17
main	27	18	1.50	much	174	150	1.16
excellent	24	16	1.50	their	265	230	1.15
information	79	53	1.49	.	4204	3714	1.13
mind	47	32	1.47	like	225	199	1.13
thanks	65	45	1.44	with	772	683	1.13
course	248	172	1.44	all	342	304	1.13
am	314	218	1.44	have	828	756	1.10
learned	33	23	1.43	to	2955	2709	1.09
				for	920	846	1.09

Note. Only normalized frequency and effect size are presented here to reveal the differences between the two types of posts. The full statistics for the log-likelihood ratio test and dispersion measure (i.e., the criteria to decide whether a word is a keyword) of each independent keyword can be found in Appendix D.

¹The normalized frequency is measured by per 100,000 words and is rounded so there is no decimal places.

²The effect size is measured by relative risk, that is the ratio of the normalized frequency in independent posts to the normalized frequency in initiating posts. The calculation is based on unrounded normalized frequency rather than the rounded one shown here.

As shown in Table 6.2, the effect size of six keywords, *joined*, *informative*, *forward*, *improve*, *keen*, *hoping*, are above 2, although their frequency (8 to 71 per 100,000 words) are relatively low compared to other keywords. All of them are used most often in the expression of intention to participate and learn in the MOOC, which will be further explained in Chapter 7. The keywords with lowest effect size are grammatical words, *have*, *to*, *for*, which are used frequently in both types of posts (range from 756 to 2955 per 100,000 words).

In summary, the keyword analysis shows that there are indeed differences at the level of word usage between initiating posts and independent posts. However, the keyword lists only reveal such difference and statistical information on the keywords, i.e., their frequency in the corpus and the effect size of their frequency comparison. To interpret these lists of keywords, and ultimately derive the linguistic features and discourse practices of initiating posts and independent posts, a follow up qualitative analysis of individual keywords based on the co-text and context they occur in is conducted next.

6.3 Interpretation of keywords: Functional grouping

Following the approach of functional grouping introduced in Chapter 4, seventeen functional groups are identified for the keywords of both types of posts, and are presented in Table 6.3. The following subsections describe ten of these groups in details by focusing on the initiating keywords, with independent keywords as reference. This is because the main aim of the thesis is on the dialogic nature of online discourse, that is those discourse practices that are likely to establish user-user interactions, thus conversations with others, as opposed to user-content interactions. Most of these initiating keywords are employed in the discourse practices for information seeking and stance-taking, as we shall see. The other seven groups of keywords are not analysed because preliminary analysis shows that they are not relevant for the discourse practices which establish user-user

interactions. Some also are used for various functions that make it hard to draw overall conclusions.

Brief discussion of these groups of keywords are presented in section 6.13.

Table 6.3 Functional grouping of the initiating keywords and independent keywords.

Functional grouping	Initiating keywords	Independent keywords
Modals/Modal expressions ¹ (6.3.1) ²	might, would, could	will, need, able ³
Hedges (6.3.2)	perhaps, seems, sort ⁴	
Evaluatives (6.3.4)	wrong, missing	easy, excellent, better, interesting, informative, great, important, good, new, difficult
Negations (6.3.5)	cannot, ca, n't	
Mental verbs (6.3.6)	wonder, wondering	aware, understanding, learned, think, agree, feel, feeling, keen, hope, hoping, looking, forward, enjoy, enjoyed, love
Communicative verbs (6.3.7)	mean, explain, tell, says, say, told, called	
Activity verbs (6.3.13)	used, tried, came	joined, affects, helps, achieve, work, gain, meet, improve
Meta-language on learning and discussion (6.3.9)	question, article	information, knowledge, course
Indefinite pronouns (6.3.10)	anybody, anyone	everyone
Polite speech-act formulae (6.3.11)	please, sorry	yes, thanks, thank
Connectors (6.3.12)	If, or, then	also, and
Quantifiers (6.3.13)	any	all, lot, much, every
Boosters (6.3.13)	surely, rather, else	really, very, definitely, always
Pronouns (6.3.13)	he	I, my, our, their
Grammatical particles (6.3.13)	the, that, there, here, does, did, was, were, 's, on, by, than	am, 'm, have, for, about, with, to, more
Punctuation (6.3.13)	,...-();?":	!.
Uncategorized (6.3.13)	example, e.g., 1, one, two ⁵ , just, numbers, why, whether, same	like, well, week, main, currently, working, opportunity, education, environment, mind

¹Modals are also grammatical, but they are categorized separately because of their crucial function in language communication and discourse analysis (Fairclough, 2003; Stubbs, 1986).

²The number in the bracket indicates the subsection in the text that describes the group.

³96% of the instances of *able* collocate with *to*, forming the modal expression *able to* (Carter & McCarthy, 2006).

⁴70% of the instances of *sort* collocate with *of*, forming the hedging expression *sort of*.

⁵*1*, *one*, and *two* arguably function as quantifiers as well, but they differed from the other quantifiers in the sense that they are numerals that specify exact amount (Biber et al, 1999) and do not have the intensifying or down-toning function in stance expression.

6.3.1 Modals/modal Expressions

Modal expressions are used by language users to modulate their attachment/detachment to the proposition they are making (Stubbs, 1986). By modulating one's proposition, the dialogic space created can be expanded or contracted accordingly (Martin & White, 2005). A clear difference is found in the usage of modal expression between initiating posts and independent posts. The modals *would*, *could*, *might* are used significantly more often in the initiating posts, compared to in the independent posts, whereas *will*, *need*, *able* are used more often in the independent posts. Based on the collocation analysis and concordance reading, the general findings are that *would*, *could*, *might* are associated with hypothetical situations and tentativeness, whereas *will* carries more certainty in contrast to its past tense form *would*, *need* refers to obligation and *able* to ability. This is consistent with their typical usage in general occasions in spoken and written language (Biber et al., 1999; Carter & McCarthy, 2006). In the elaboration below, *would* is used as an example, with occasional comparison to *could* and *might*, to relate the meaning of hypothesis and tentativeness to discourse practices of information/help seeking and stance-taking in initiating posts that start user-user interactions.

Of the 9174 instances of *would* in the initiating posts, the five most frequent bigrams⁹ are: *I would* (n=2295), *it would* (n=1052), *they would* (n=268), *this would* (n=268), *that would* (n=242)¹⁰. Similar patterns are found for *might* (n=1806) and *could* (n=4447), with *I might* (n=180), *it might* (n=181), *I could* (n=786) and *it could* (n=270) as the most frequent bigrams. When these modals are used with personal pronouns, they typically refers to one's subjectivity, whereas when used with *it*, *this*, *that* and other inanimate objects, they typically refer to a prediction of an event by minimizing one's agency but implying objectivity (Biber et al., 1999; Du Bois, 2007). This differentiation between

⁹ A sequence of two words.

¹⁰ It is worth noting that frequency of words or linguistic formal structures in human language follows Zipf's law such that their frequency of usage roughly follows a logarithmic scale (Evert, 2008), as can be seen in the frequency distribution of these top five phrases where there is a sharp drop of frequency from the most used phrase to the second most used, and the next etc.

self-attribution and distancing oneself is investigated by examining the two most frequent phrases *I would*, *it would* and a random sample¹¹ of other bigrams.

6.3.1.1 *I would*: Expressing interest or seeking information

The most frequent phrase for self-attribution is *I would like* (n=490), in which *like* is also an initiating keyword. This phrase, and semantically similar phrases, *I would love* (n=83), *I would be interested* (n=54), are typically used in the initiating posts to seek information by expressing an interest to know (e.g., “...*I would like* to know how it works the technique...¹²”; “...*I would be interested* to delve deeper. [...] ¹³ any recommendations for books or sites [...] further?”¹⁴). Posts consisting of this practice of expressing interest typically receive encouraging replies from facilitators and other users, though the conversations usually stop after one reply to the request and can be short-lived, as shown in Figure 6.1.

¹¹ The random sampling started with 30 comments containing the keyword, then continued for another 30 samples until no new patterns were found. Effort was also made to ensure comments from all 12 MOOCs were sampled.

¹² <https://www.futurelearn.com/courses/soils/1/comments/6359171>

¹³ [...] indicates omission of some of the quotes.

¹⁴ <https://www.futurelearn.com/courses/soils/1/comments/6435167>

Figure 6.1 Thread 5157379.

Thread 5157379
Source: <https://www.futurelearn.com/courses/dyslexia/1/comments/5157379>

Initiating post 2015-05-17 08:21:48 Like: 0
User d1-358
Very interesting activities but not always doable with high school students. As I said before, *I would like* to use a software where students could virtually manipulate letters, listen to different sounds and working autonomously to improve their phonological awareness. I don't know if a software like this exists (I hope so!) but I think this would save time and be more fun for teenagers.

Reply 1 2015-05-17 08:44:03 Like: 1
Facilitator
IT's not a software but it might help:
<http://www.bbc.co.uk/worldservice/learningenglish/grammar/pron/>

Note. Threads will be presented this way throughout the thesis, with the URL linked to the FutureLearn platform where the thread can be found as source. The time of posting and “like” received are presented behind the label of each comment.

In the initiating post of Figure 6.1, the user “*would like* to use a software” but does not “know if a software like this exists”, and at the same time expresses their stance that this particular kind of software “*would* save time and be more fun for teenagers”¹⁵. There is no explicit request for the software, and in fact the user recognizes that it might not exist. Yet, the facilitator responds with relevant information. This suggests that expression of interest can create a dialogic relationship with potential readers. However, this relationship is often short-lived and limited to a question and answer routine, which is found to be typical of MOOC discussions (Poquet et al., 2018). In contrast to *I would* to express interest to know, *I could* is used to express one’s incapability (e.g., “...with Internet explorer *I could not* upload the picture...”¹⁶) to achieve information-seeking. Expressing an “unknowing” status is another way of establishing a dialogic relationship with the other who has a

¹⁵ Present tense, instead of past tense is used in thesis for describing the analysis and interpretation of users’ discourse.

¹⁶ <https://www.futurelearn.com/courses/exploring-our-oceans/1/comments/279986>

“knowing” status (Heritage, 2012), and can also be achieved with the initiating keywords *wrong* and *missing*, as will be explored later.

6.3.1.2 *I would*: Positioning oneself

Besides seeking information, *I would like/love* is also used for positioning oneself in relation to what has been raised in the context (Du Bois, 2007). Users use *I would like/love* to express their positioning regarding what has been mentioned in the course (“...*I would like* to see more incentive-driven remuneration to suppliers to the public sector...¹⁷”). Stance-taking is also realized when users use communicative verbs and mental verbs with *I would*, including *say* (n=101, which is also an initiating keyword), *think* (n=42), *expect* (n=36), *imagine* (n=31), *suggest* (n=24), *prefer* (n=21), *guess* (n=19). These verbs are used for expressing one’s subjectivity, yet *would* helps soften or hedge this subjectivity (Carter & McCarthy, 2006; McDonald & Woodward-Kron, 2016). This explicit attribution to oneself by *I* and communicative verbs, and the tentativeness as introduced by *would*, frame the initiating post as stance-taking with epistemic uncertainty, rather than a matter of fact.

Similarly, *I might* and *I could* are also used for positioning in stance-taking, although *I might* often carries the meaning of personal intention (e.g., “...Interested to see how *I might* improve my garden without having to dig it all up again....”¹⁸; “...I think, like Marcus Aurelius, *I might* just adjust the dosage according to my need at the time!”¹⁹) and *I could* is used to recount personal stories carrying the meaning of ability (e.g., “...I use a side plate for my main meal and for dessert the smallest of the smallest bowl that *I could* find...”²⁰; “...I can hardly imagine how *I could* implement them as a common practice...”²¹). Given that they are taking a stance regarding the course content, it is possible for others to either align or disalign with them in this shared context, thus creating a dialogic space, as shown in Figure 6.2.

¹⁷ <https://www.futurelearn.com/courses/contract-management/4/comments/18272032>

¹⁸ <https://www.futurelearn.com/courses/soils/1/comments/6184384>

¹⁹ <https://www.futurelearn.com/courses/ancient-health/1/comments/20083364>

²⁰ <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/19598537>

²¹ <https://www.futurelearn.com/courses/dyslexia/1/comments/5098328>

Figure 6.2 Thread 4954168.

Thread 4954168

Source: <https://www.futurelearn.com/courses/dyslexia/1/comments/4954168>

Initiating Post 2015-05-12 11:58:29 Like: 1

User d1-4010

I liked most of these activities - though since we have been told that for dyslexic students we need to provide explicit explanations of rules, I assumed they must either be revision games or games that would then lead into a discussion about spelling rules? I think I might change the domino game in terms of making the students match words that have the same 'ch' beginning or 'ch', 'tch' ending rather than just the same word. Perhaps we could then look at why they have the same onset / rime. I loved the slides - have used something v similar as a primary teacher teaching phonics and love the fact that its very kinasthetic. I wasn't so sure about the odd one out activity, although I liked it I could see that it was quite difficult and would maybe use it in groups or as a team game on the whiteboard rather than an individual activity. Using it as a team game could make it more multi sensory as well, as in its current state it relies on students having a very good visual perception.

Reply 1 2015-05-12 14:56:06 Like: 1

User d1-2591

I liked your adaptation for the tch/ch sound Louise, to make it more challenging but then as I don't teach sts with dyslexia on a regular basis, perhaps that would make it too challenging? It would certainly be unchallenging for an non-dyslexic student and that creates a problem of using the same materials in a mixed ability class. We might need to use different versions with different students, depending on their needs.

Note. Underlined emphasis is mine to highlight the alignment of reply in relation to the initiating post.

In Figure 6.2, the initiating post is in response to a discussion prompt “How could you adapt and implement these activities and teaching aids to better suit your teaching context?”. The initiating user takes a position with “I *might* change the domino game”, which indicates a user-content interaction with the learning materials. This positioning also sets up a potential dialogic space for other users to align or disalign with, as shown in the reply. The replying user seems to align with “I liked your adaptation” but disalign at the same time by suggesting it “*would* make it too challenging”. This example shows that an initiating post can start user-user interaction as well as engaging in user-content interaction.

A different stance-taking practice can be found for *could* and *might*. They are also used for admitting one's "unknowing" status (Heritage, 2012) by making explicit the possibility of one being wrong (e.g., "...*I could be* wrong here as I have not read Celsus..."²²; "...I think others have usefully commented - though perhaps - and *I might* have missed it - we are looking at the..."²³). The admission that one might be wrong hedges one's stance and can be a more explicit invitation for alternative viewpoints compared to *I would* + communicative and mental verbs. This admission can also establish a dialogue with a potential audience who does not agree.

6.3.1.3 *it would be* + adjective: Evaluation as an introductory frame to stance or interest

It would is most frequently used in the form of *it would be* + adjective²⁴ (n=356), including *interesting* (n=108), *useful* (n=34), *great* (n=33), *nice* (n=33), *good* (n=28). The phrase *it would be* + adjective is used to voice evaluation of what follows. However, this evaluation is not the same as evaluation without modals, such as "It is great!". The evaluation that is modified by *would* can serve as an introductory frame for the users' hypothesis (e.g., "...*It would be unsurprising* if blind people in the ancient world were encouraged to fill the role of performing singer/poet."²⁵), idea (e.g., "...*It would also be interesting* to see if the discourse on RASIM has changed over time..."²⁶), or information-seeking (e.g., "...so *it would be nice* to know if they can actually participate in soil pollutants mobility..."²⁷). A similar construction can be found with *it might be* + adjective which is used to express politeness in suggestions (e.g., "...*It might be* better if the link to the NERC planning page opened in another window..."²⁸). This framing has the function of distancing oneself from the hypothesis or ideas proposed, given the use of *it*, in contrast to *I would* above (Biber et al., 1999; Hyland, 2005).

²² <https://www.futurelearn.com/courses/ancient-health/1/comments/19460541>.

²³ <https://www.futurelearn.com/courses/ancient-health/1/comments/21016703>

²⁴ Based on automatic classification of adjective by Treetagger (Schmid, 1994).

²⁵ <https://www.futurelearn.com/courses/ancient-health/1/comments/19676245>

²⁶ <https://www.futurelearn.com/courses/corpus-linguistics/1/comments/334805>

²⁷ <https://www.futurelearn.com/courses/soils/1/comments/6326168>

²⁸ <https://www.futurelearn.com/courses/exploring-our-oceans/1/comments/447304>

6.3.1.4 *it would* and other instances of *would*: Prediction with uncertainty

Besides being used in the frame of evaluation, *it would* is also used to speculate in conditionals (e.g.,

“...If inaccessible products were boycotted on this basis *it would* certainly draw the attention of the organisations...”²⁹). This function of speculation is common in other instances of *would*. An analysis of a random sample reveals that users use *would* to make tentative claims (e.g., “I think the main ingredients in a living soil *would* be a combination of sand silt clay, organic matter...”³⁰), construct hypothetical situations (e.g., “...Had all the stakeholders been involved from the onset in the planning and coordination, this incident *would* have not occurred...”³¹), speculate (e.g., “...I have found some wiki information that indicates a lower melting point for ilmenite. So this *would* be consistent with an upper limit for the impact induced temperature rise...”³²), or express doubts or questions (e.g., “...Now I'm wondering if it is just herbivore droppings that they process, or *would* any poo do for them?...”³³). Similar patterns and functions can be found for *it might*, *it could* and other instances of *might* and *could* that do not follow *I* and *it*. In all these occasions, *would*, *could* and *might* have the function of making tentative statements rather than facts and assertions. Tentative proposition is similar to the practice of plausible reasoning in research articles in which knowledge is constructed as tentative and subjective to debate and future investigation, thus creating a dialogue space (Hyland, 2005; W. Yang, 2014).

6.3.2 Hedges

Hedges are used by language users to downplay their assertions or withhold their commitment to propositions so that they do not appear to be unduly authoritative or sound matter-of-fact (Biber et al., 1999; Carter & McCarthy, 2006). No hedges are found to be significantly frequent in independent posts. Three hedges, *perhaps*, *seems*, and *sort of* are found significantly used in initiating posts.

²⁹ <https://www.futurelearn.com/courses/digital-accessibility/2/comments/19694444>

³⁰ <https://www.futurelearn.com/courses/soils/1/comments/6180604>

³¹ <https://www.futurelearn.com/courses/contract-management/4/comments/18035596>

³² <https://www.futurelearn.com/courses/moons/1/comments/649513>

³³ <https://www.futurelearn.com/courses/soils/1/comments/6268861>

Based on the collocation analysis and concordance reading, the general finding is that these hedges make explicit users' epistemic uncertainty and "unknowing" status regarding their propositions, and may appeal to other knowledgeable users to pitch in. This practice of stance expression is elaborated below, using *perhaps* and *seems* as examples.

6.3.2.1 *perhaps*: Softening stance

In initiating posts, *perhaps* (n=1060) is used to soften a suggestion imposed on the other users (e.g., "*Perhaps* for possessing more information you should interact with more interested parts..."³⁴), and to provide possible explanation or interpretation (e.g., "...*Perhaps* the baby feeders of ancient times were likewise silent killers."³⁵; "...This could *perhaps* mean that British fiction, in comparison with American fiction, uses more dialogue?"³⁶). Another function of *perhaps* found in the initiating posts is when users take a step back when raising their claims or questions (e.g., "...*Perhaps* this is due to my lack of practice in reading essays like those!"³⁷). This is similar to the admission of one's "unknowing" status as found in *could* and *might* explained earlier.

6.3.2.2 *seems*: An introductory frame to stance

Out of 1588 *seems*³⁸ found in the initiating posts, the most frequent form is *it seems* (n=658), *there seems* (n=70) and *this seems* (n=65). When *it seems* is followed by *to me* (n=119), either in *it seems to me* or *it seems + noun/adjective phrase + to me*, the phrase introduces personal perceptions (e.g., "Ancient medicine *it seems to me* defined itself at two levels..."³⁹) or attitudes (e.g., "...*it seems to me* incredible that basic healthy foodstuffs are often expensive so that..."⁴⁰). In other occasions, *it seems*

³⁴ <https://www.futurelearn.com/courses/contract-management/4/comments/18185697>

³⁵ <https://www.futurelearn.com/courses/ancient-health/1/comments/20036129>

³⁶ <https://www.futurelearn.com/courses/corpus-linguistics/1/comments/376236>

³⁷ <https://www.futurelearn.com/courses/corpus-linguistics/1/comments/375560>

³⁸ The keyword analysis reveals that *seems* is a keyword but other forms such as *seem* and *seemed* are not. This could be due to the observation that "seem" is used with inanimate subject (*it*, *there* and *this*) more often.

³⁹ <https://www.futurelearn.com/courses/ancient-health/1/comments/20284559>

⁴⁰ <https://www.futurelearn.com/courses/ancient-health/1/comments/20572816>

is used without being followed by *to me* to make impersonal claims (e.g., “...*It seems* that they had to do the work themselves...⁴¹”). The introduction of stance with *it seems* is similar to *it would be +adjective* discussed previously.

In some occasions, the stance introduced by *seems* is observed to be challenging a presumed perception. For example, “...*it seems easier* for the children with SpLDs to acquire reading and writing. However, when it comes to English language learning, which is a language with less 'transparent orthography' the same learners have to struggle much harder...”⁴²; “...Everyone *seems* to think there ought to be differences in pay, but no one talks about what those pay differences mean...”⁴³. In both examples, *seems* is used to introduce a presumed acceptable view, before the users challenge this view, thus setting up a dialogic space with opposing voices.

6.3.3 Concluding remarks on modals and hedges

In this analysis, modals and hedges are grouped separately. However, out of the 2686 initiating posts that contain at least one of the three key hedges, 1214 (45%) of them also contain at least one of the three key modals, suggesting that users sometimes make use of both modals and hedges to express their stance in the initiating posts. In general, both groups of keywords function similarly in stance-taking, i.e., framing a proposition as hypothesis, possibility and uncertainty, rather than making a bare assertion. This softens one's arguments and acknowledges alternative possibilities that other users may have, thus opening up a dialogic space for others to challenge or expand the original stance, or even raise alternative views (Martin & White, 2005). Similar conclusions have been made by Drasovean & Tagg (2015) on TED commenting spaces, although they do not specify if these discourse practices are in posts that receive replies. Lastly, both hedges and modals are also used to explicitly declare one's 'unknowing' status or interest to know to establish a dialogic relationship with those who know.

⁴¹ <https://www.futurelearn.com/courses/dyslexia/1/comments/5038651>

⁴² <https://www.futurelearn.com/courses/dyslexia/1/comments/4582700>

⁴³ <https://www.futurelearn.com/courses/inequalities-in-personal-finance/1/comments/4348386>

6.3.4 Negative evaluative keywords: Admitting one's mistakes

Wrong (n= 589) and *missing* (n=266) are two initiating keywords that can potentially be categorized as evaluative and negative in nature (e.g., "...It's naive and *wrong* to refer to LDL as 'bad cholesterol'..."⁴⁴; "...there is a clear policy directive to do something about housing crisis that is sadly *missing* in the UK at present..."⁴⁵). In contrast, among the independent keywords, *easy*, *excellent*, *better*, *interesting*, *informative*, *great*, *important*, *good*, *new*, *difficult*, that can be classified as evaluative, the majority are positive in nature. These evaluative keywords are mainly used in the independent posts by users when they comment positively on the course materials.

Intriguingly, unlike Biber et al. (1999) who only categorized *wrong* as negatively evaluative, the analysis of *wrong* in the initiating posts revealed a notable function – admitting one's mistake. This admission is used to make concessions to one's stance (e.g., "...It has a clastic appearance, but I may be *wrong*..."⁴⁶), or to explicitly declare "unknowing" status for seeking help (e.g., "...I really dont know what I am *doing wrong*..."⁴⁷; I couldn't get the excel paste working correctly, all appeared in one column - must be doing something *wrong*?...⁴⁸). Similar trends can be noted about *missing* (e.g., "Maybe we could try and summarise Chomsky's criticism of CL? Early CL:1 - representativeness - addressed; [.....] I am sure I am *missing* something?"⁴⁹; If this is correct why doesn't most of the heat simply radiate into space, or am I *missing* something?⁵⁰).

In summary, although *wrong* and *missing* generally carry evaluative meaning, a closer look at the keywords in context reveals one notable function that has not been documented before (e.g. by Biber, et al., 1999), that is, admitting one's possible mistake. This admission is used not to negatively evaluate anything but to hedge one's claim, similar to the function of modals and hedges discussed in section 6.3.1 and 6.3.2, in turn establishing a dialogic relationship with others who know or have

⁴⁴ <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/18952923>

⁴⁵ <https://www.futurelearn.com/courses/inequalities-in-personal-finance/1/comments/4420945>

⁴⁶ <https://www.futurelearn.com/courses/moons/1/comments/659777>

⁴⁷ <https://www.futurelearn.com/courses/learn-to-code/1/comments/8493057>

⁴⁸ <https://www.futurelearn.com/courses/corpus-linguistics/1/comments/299144>

⁴⁹ <https://www.futurelearn.com/courses/corpus-linguistics/1/comments/266243>

⁵⁰ <https://www.futurelearn.com/courses/moons/1/comments/525007>

alternative views. These findings also highlight the importance of analysing the keywords in context when interpreting the results of keyword analysis.

6.3.5 Negation

Negation contradicts a proposition, and is typically expressed by the negative particle *not* and *n't* (Biber et al., 1999). Although negation is hypothesized as contracting dialogic space (Martin & White, 2005), in the keyword analysis, *n't* (n=8299), *can't* (n=1139), and *cannot* (n=731) were found to be used significantly more often in initiating posts than in independent posts. There are no independent keywords indicative of negation. Collocation analysis reveals that *I* is one of the top ten 5-word left-collocates for these three keywords, with *I* as collocate for *n't* (n=3455), *can't* (n=553), *cannot* (n=250). The pattern *I...n't/can't/cannot* suggests negation acts by users. In contrast, other patterns, such as *it ... n't* (n=429), *we ... n't* (n=390), *they ... n't* (n=359), *we cannot/can't* (n=142), *they cannot/can't* (n=86) are typically used for expressing negative propositional content. *you cannot/can't* (n=101) can be negation acts targeting the addressee, or used in negative propositional content when *you* is referred to people in general. This differentiation of negation acts and negative propositional contents are based on Biber et al. (1999).

Negation acts are elaborated on in detail in the following subsections because, as we shall see, they are similar to disalignment or disagreement that is indicative of interactivity (Baym, 1996; Kleinke, 2010; Tanskanen, 2007). Negation acts are sometimes carried out by users with phrases such as *I don't think* (n=298), *I don't know* (n=283), *I don't have* (n=123), *I didn't know* (n=113), *I cannot/can't see* (n=89). They are explored below in terms of their function for information seeking and stance-taking in initiating posts.

6.3.5.1 *I don't know*: Information/help seeking

A major function of negation acts is to indicate one's inability or "unknowing" status and thus call for help from others, as shown in the phrases *I don't know* (e.g., "...*I don't know* if this is the kind of a

priori decision that can be tested using corpus linguistics?...⁵¹), *I don't understand* (e.g., "*I don't understand* how Pluto can revolve around nothing. If there is nothing there to provide the gravitational force to keep it in orbit why does it stay there?"⁵²), and *I don't get* (e.g., "...When I try to reproduce them with AntConc / LOB_tagged, *I don't get* any results!..."⁵³). In these occasions, the replies received are answers to their questions or requests, although the interactions can be short-lived. This function is also achieved by using *could not/n't*, and admitting being *wrong* or *missing something*, as discussed earlier.

6.3.5.2 *I don't think/understand*: Stance expression

When negation acts are used for expressing stance, they can be framed in the form of disalignment or disagreement, such as in *I don't think*. However, negation acts framed in epistemic uncertainty, such as *I don't understand*, can also be used to take stance but with reduced force. These two ways of stance-taking are explored in Figure 6.3 and Figure 6.4 respectively to show that negation does not necessarily deny what has been said, but sets up two voices in the current communicative context (Bakhtin, 1981; Jordan, 1998), at least one of which is the user's voice. The stance can be voiced in response to the discussion prompt or course content, thus indicating user-content interaction. Figure 6.3 illustrates an initiating post which is in response to a discussion prompt.

⁵¹ <https://www.futurelearn.com/courses/corpus-linguistics/1/comments/307032>

⁵² <https://www.futurelearn.com/courses/moons/1/comments/536555>

⁵³ <https://www.futurelearn.com/courses/corpus-linguistics/1/comments/537728>

Figure 6.3 Thread 18994217

Thread: 18994217

Source: <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/18994217>

Discussion Prompt: Do you think that obesity is a 'disease'?

Initiating Post 2017-01-20 11:04:23 Like: 12

User n4-3207

I don't think obesity is a disease, but it is a condition that can lead to chronic illnesses, and is both caused by personal choices and an environment where processed food is loaded with often hidden sugar.

Reply 1 2017-01-20 11:09:22 Like: 0

User n4-2026

I agree

Reply 2 2017-01-20 11:19:24 Like: 0

User n4-2561

I agree

Reply 3 2017-01-20 13:05:57 Like: 0

User n4-3410

I agree with you, very interesting thoughts!

In Figure 6.3 , the negation, *I don't think ...but* can be said to introduce a differing opinion, thus possibly expanding the dialogic space in the discussion. However, in this thread, the initiating post only triggers a series of agreements rather than a discussion with more content. Besides *I don't think*, several other phrases with perceptive verbs following *n't*, *can't* and *cannot* are also used in initiating posts by users to introduce their stance, such as *see* (e.g., "...houses are overpriced in UK, especially London and *I can't see* any housing bubble happen soon..."⁵⁴; "...if they choose to remain here, and *I don't see* that we can say that they cannot...."⁵⁵), *imagine* (e.g., "...*I cannot imagine* living in one of these 'streets in the sky' which used to be the euphemism for this kind of high density

⁵⁴ <https://www.futurelearn.com/courses/inequalities-in-personal-finance/1/comments/4413310>

⁵⁵ <https://www.futurelearn.com/courses/inequalities-in-personal-finance/1/comments/4365821>

living ..."⁵⁶). This negation act with / turns the stances into a personal opinion rather than a matter of fact.

Besides responding to discussion prompts, negation acts can sometimes be in response to the course content, as shown in Figure 6.4.

⁵⁶ <https://www.futurelearn.com/courses/inequalities-in-personal-finance/1/comments/4442776>

Figure 6.4 Thread 19605888

Step content:

Sugars in the diet

Dietary sugar is an integral part of our western diets. In its natural form it can be found in most fruit and vegetables but it can also be added to number of processed food and drinks.

In this video, Geraldine McNeil of the University of Aberdeen explains the different types of sugars available in the diet, and gives a brief overview of how sugar consumption has changed over the last century. She explains terms such as ‘free sugars’ and discusses the hype surrounding our current sugar consumption levels and health. She gives examples on how much sugar is contained in a number of commonly consumed foods, and tells us how much we should consume for optimal health.

Thread: 19605888

Source: <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/19605888>

Initiating Post 2017-02-12 09:31:34 Like: 1

User n4-1568

Interesting video, I somehow *don't* understand though how in the UK the soft drinks get the blame for the excessive intake of sugar as when you go to the supermarket most drinks are sugar free, finding squash without artificial sweeteners in is actually almost impossible. Then when you look at sweets, sauces, ready made meals, yogurt pots, even baby products, they're all full of sugar!

Reply 1 2017-02-14 05:42:12 Like: 1

User n4-3647

Yeah, I also have a doubt about sugar free food/gum. Maybe they are only free sugars-free, not table sugar free?

Reply 2 2017-02-14 11:48:01 Like: 0

User n4-1568

Exactly. I don't think they are a healthier alternative.

In Figure 6.4, *I don't understand* is used to introduce a stance that “most drinks are sugar free” but many foods are not in response to the “Interesting video”. This attracts alignment from another user, “Yeah, I also have a doubt about sugar free food/gum”. At the same time, this replying user seems to also interpret the initiator’s “I somehow *don't* understand though how in the UK the

soft drinks get the blame” as a question, to which they reply with “Maybe they are only free sugars-free”. However, from the reassertion of the initiator in reply 2 “Exactly, I don’t think ...”, it could be argued that the “I don’t understand” in the initiating post is more likely to function as weakening the force of the stance by indicating their own epistemic uncertainty, in comparison to *I don’t think, I don’t see* which is more explicit in their negation. Similar examples can be found for *I don’t know* (e.g., “...Also Medusa (who was very badly treated) was given the curse of turning anything to stone when she looked at it. *I don’t know* how significant that is though...”⁵⁷).

In these situations, rather than the literal meaning of not having sufficient knowledge which is used in information seeking, *I don’t understand/know* seems to be used as a way of making concession when users raise their ideas but display less than full epistemic commitment by stating their uncertainty to avoid expressing an unequivocal stance (Tsui, 2009; Weatherall, 2011). Specifically, this usage of *I don’t know* in the initiating posts, i.e., start of a conversation thread, differs from replying to a query in a conversation. It is possible that *I don’t know*, as well as *I don’t understand*, in the initiating posts may have the function of opening up a conversation floor because the admission of lack of knowledge could invite divergent contributions from other participants (Grant, 2010).

6.3.5.3 Concluding remarks on negation

In-depth analysis of negation acts with the pattern *I...n’t* reveals that taking opposing stance does not necessarily contract a dialogic space. Stance is expressed indirectly with an “unknowing” status such as *I don’t know/understand*. This expression of “unknowing status” is also used when user request help. These functions attest to the possibility that negations, when not embedded in the propositions, are sometimes used to make concession and avoid commitment to one’s proposition while raising a claim or request (Tsui, 2009; Weatherall, 2011), thus opening up a dialogic space rather than closing it down.

⁵⁷ <https://www.futurelearn.com/courses/ancient-health/1/comments/19678314>

However, it is worth noting that almost half of the time, *n't*, *can't* and *cannot* are used in patterns without *I* for stating a negative proposition (e.g., "...the Geological community *cannot* find any evidence to support such a division in the geological record...."⁵⁸). The negative proposition in the context without *I* in the initiating posts could be more assertive as it is not personalized as one's opinion, unlike the negation act of *I don't* + perceptive verbs which weakens the force of denying.

The findings suggest that denying or disclaiming do not necessarily contract but expand dialogic space (Martin & White, 2005). This happens when users take opposing stance by expressing epistemic uncertainty and personalizing the stance as one's opinion rather than matter of fact. Although the initiating posts do not address a specific person, this finding is consistent with Baym's (1996) investigation on Usenet users' disagreement towards others, which shows that mitigated and qualified disagreement leaves room for others to continue conversations.

6.3.6 Mental verbs: *wonder* and *wondering* to hedge questions and stance

A mental verb refers to mental state or activity experienced by humans but not necessarily observed by others (Biber et al., 1999). A clear difference is found between initiating posts and independent posts in the usage of mental verbs. Only two mental verbs are found significantly used more often in initiating posts, whereas 15 mental verbs are found significantly used more often in independent posts.

The two mental verbs found to be the initiating keywords are *wonder* (n= 990) and *wondering* (n= 338) that carry the meaning of uncertainty in epistemic status and normally used in indirect questions. Collocation analysis shows that *I* is the most frequent collocate on the 5-word window on the left of *wonder* and *wondering*, whereas on the right, *if*, *how*, *whether*, *what*, *why* are among the top ten 5-word collocates. This pattern suggests that users raise indirect questions by using *I wonder* or *I was/am/a'm wondering* (e.g., "*I wonder* if there is a relationship between

⁵⁸ <https://www.futurelearn.com/courses/soils/1/comments/6457187>

memorizing and dyslexia”⁵⁹). In a corpus findings on conference calls, Camiciottoli (2009) also finds that indirect requests function to elicit more information than direct requests. The uncertainty in epistemic status as expressed by *wonder* and *wondering* help opening up a space for others to pitch in, although sometimes the replies can be sharing similar uncertainty.

However, a close reading of initiating posts containing *wonder* or *wondering* reveals that some instances may not simply be raising an indirect question but also voicing one’s opinion. This interpretation can be deduced from what follows the questions framed by *wonder(ing)* in the initiating posts and the replies received, as shown in Figure 6.5.

Figure 6.5 Thread 4329045

Thread 4329045

Source: <https://www.futurelearn.com/courses/inequalities-in-personal-finance/1/comments/4329045>

Initiating Post 2015-03-26 23:21:19 Like: 1

User f1-265

I *wonder* what the results would look like if we took the money spent by the poorest 40% on cellphones, flat-screen TVs, computer games, NetFlix, overseas holidays, cars, cigarettes, alcohol, etc. and added that money to their wealth? I say this because that money is a much greater % of income for the poor than it is for the rich.

Reply 1 2015-03-27 13:10:44 Like: 1

User f1-530

I'm not entirely sure what point you're making here, but if it's a judgement that the poorest 40 % shouldn't be spending money on these things, I don't think it's as black and white a case as you may be implying. Yes, it's important to live within one's means, but those struggling

[There are 10 more replies afterwards]

In Figure 6.5, the user first asks “I *wonder* what the results would look like...” but answers themselves immediately “I say this because...”, suggesting that this is a rhetorical question for

⁵⁹ <https://www.futurelearn.com/courses/dyslexia/1/comments/4527268>

stance-taking. Furthermore, the reply first received is a disalignment, “I don’t think it’s as black and white a case as you may be implying” rather than an answer to a question, suggesting that other users interpret this “*I wonder*” as stance-taking. As *I wonder* carries uncertainty, using it to introduce stance may also have a hedging effect, similar to the admission of being *wrong* or *I don’t know*, as explained earlier.

Therefore, the mental verbs *wonder* and *wondering* found in the initiating posts are typically used to raise indirect questions or to hedge stances. This contrasts with the 15 mental verbs among the independent keywords. Of these, *think*, *aware*, *understanding*, *learned* carry cognitive meaning, especially related to learning, probably because the online discussion is in a learning setting. Of the others, *agree*, *feel*, *feeling*, *keen*, *hope*, *hoping*, *looking forward*, *enjoy*, *enjoyed*, *love* carry emotional meanings expressing various attitudes or desires, of which almost all are positive in nature, corresponding to the other positive evaluative independent keywords. The discourse function of these independent keywords will be discussed in the next chapter, where *think* and *agree* that are commonly used in stance-taking and social interactions (Baym, 1996; Kärkkäinen, 2003; Pomerantz, 1984), will be closely examined.

6.3.7 Communicative verb forms

A communicative verb refers to the action of transmitting or receiving information (Biber et al., 1999). Five communicative verb forms, *say* (n=1476), *says* (n=387), *tell* (n=480), *told* (n=440), *called* (n=521), *explain* (n=356), *mean* (n=752), are found to be used significantly more frequently in initiating posts than in independent posts, whereas no communicative verbs are found among the independent keywords. *Say*, *tell* and *mean* are usually used in stance expression, *says*, *told* and *called* are used for intertextuality to other sources for evidentiality in stance expression, *tell*, *mean* and *explain* for information/help seeking, as explained below.

6.3.7.1 *say, tell and mean*: Stance expression

When used for stance expression, *say* is used quite frequently in the form of *I would say*, as discussed in section 6.3.1 (e.g., “I would *say* that we don't value the planet's resources properly at all.....”⁶⁰), *have to say* (e.g., “Unfortunately once again I *have to say* that eating healthy in my country it is very expensive.....”⁶¹), *evaluative + to say* (e.g., “.....*Hard to say* if they are hygienic by our standards.....”⁶²). Similar patterns can be found for *tell* (e.g., “.....I can *tell* you that much like the UK the US has 75% of people who are working class or below.....”⁶³). This stance expression involves an explicit mention of a personal opinion with modalization, rather than bare assertion.

In contrast to *say* and *tell*, *mean* is generally used to reformulate what has been mentioned previously either in the course content or presumptions, into questions (e.g., “Does that *mean* that processes that take place on the surface of the Earth have similar processes deep on the ocean bed?? If so, then cant we predict tsunamis ???”⁶⁴), negations (e.g., “.....There are many other manual jobs for which a retirement age in the late 60s is totally unreasonable. But that does not *mean* that they cannot contribute to society.....”⁶⁵), or modalized expressions (e.g., “.....To my knowledge with online purchases, as well as, with other contracts, the buyer's country is always the place of jurisdiction, automatically. This would *mean* if one bought from abroad the local jurisdiction has some kind of process of dealing with a claim in place”⁶⁶). These instances of reformulation show that users are acknowledging general viewpoints or what has been discussed. At the same time, they develop their comments on what has been mentioned. In the reformulation, their stance is raised against a common ground, thus setting up a shared dialogic space. The dialogic space is also expanded due to their reformulation.

⁶⁰ <https://www.futurelearn.com/courses/soils/1/comments/6471416>

⁶¹ <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/19266631>

⁶² <https://www.futurelearn.com/courses/ancient-health/1/comments/20412992>

⁶³ <https://www.futurelearn.com/courses/inequalities-in-personal-finance/1/comments/4304515>

⁶⁴ <https://www.futurelearn.com/courses/exploring-our-oceans/1/comments/299332>

⁶⁵ <https://www.futurelearn.com/courses/inequalities-in-personal-finance/1/comments/4371338>

⁶⁶ <https://www.futurelearn.com/courses/contract-management/4/comments/17892710>

Another interesting observation is the discourse marker *I mean*, which accounts for one fourth of the bigrams containing *mean* in the initiating posts. Consistent with previous findings (Carter & McCarthy, 2006; Fung & Carter, 2007), *I mean* in the initiating posts acts as a discourse particle that does not contribute to propositional meaning but functions to connect between utterances and indicate interactions with others. Similar to its functions in classroom or daily conversations (Biber, Conrad, & Cortes, 2004), it is typically used to introduce elaborations, expansion or clarifications after the users raise their opinions (e.g., “I truly believe that food addiction exists. Certain foods release endorphins in the brain. Food to me is a substance. *I mean* think about how we respond to chocolate?”⁶⁷) or questions (e.g., “From what I gather, we should teach them the symbols as well? I mean the IPA-based pronunciation system? Won't that be too much? Or do we just mention and not focus on them?”⁶⁸). This usage of *I mean* could be said to be interactive and broadening the dialogic space because users are deepening or clarifying their own arguments as if they were responding to an imaginary audience’s request to clarify (Bakhtin, 1981).

6.3.7.2 *says, told* and *called*: Intertextuality for evidentiality and information sharing

Called, told, says are used to refer to external sources as evidence for stance. *called* is used mainly to introduce terminology or some entities (e.g., “...I was reminded of a short documentary film *called* “Natural World: A farm for the future...”⁶⁹; “...the patient has requested to die at home surrounded by memories of their life and loved ones we have a system *called* FAST TRACK or RAPID DISCHARGE...”⁷⁰). *Told* and *says* can be indicative of hear-say evidentiality (e.g., “...I've been *told* that best to avoid (and counteract) cancer is to ...”⁷¹; “...Interesting that this issue has been raised again this week and Peter Hain *says* that there has been enough discussion”⁷²). The intertextuality as

⁶⁷ <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/19049804>

⁶⁸ <https://www.futurelearn.com/courses/dyslexia/1/comments/4934562>

⁶⁹ <https://www.futurelearn.com/courses/soils/1/comments/6245068>

⁷⁰ <https://www.futurelearn.com/courses/palliative/1/comments/17228088>

⁷¹ <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/19342663>

⁷² <https://www.futurelearn.com/courses/exploring-our-oceans/1/comments/373202>

realized by these practices does not only present evidence for supporting a stance, but also expands the dialogic space because it is pointing to other sources which could also form the common ground for other users to comment on. Furthermore, this intertextuality to other sources can also be considered information sharing since other users can gather more information from the sources.

6.3.7.3 *tell, mean and explain*: Information/help seeking

Both keywords *explain* and *tell* are typically used in information-seeking questions (e.g., “Can I just ask can anyone *explain* what internal geological activity means...?”⁷³; “Could someone *tell* me if there is any significant design distinctions between these wetland features and a swale?”⁷⁴). This is based on the observation that words such as *someone, somebody, please, can, could, anyone* all feature in the top ten 5-word window left collocates of *explain* and *tell*, while *what* and *how* in the top ten 5-word window right collocates of both words, whereas *why* features as the top right collocate of *explain*. In these occasions, there is an explicit call for others to *tell* or *explain*. In contrast, when *mean* is used, it is typically used in a direct question in the patterns *does this/that/it/mean, do/did they/we/you mean*.

6.3.8 Concluding remarks on mental verbs and communicative verbs

The only mental verbs found to be significantly used in the initiating posts are *wonder* and *wondering*, which are used for stance expression and information seeking. In contrast, the occurrence of several mental verbs conveying positive emotion or attitudes in the independent posts suggests explicit appreciation to the course, i.e., user-content interaction. As would be expected of user-user interaction, communicative verbs are found to be significantly more frequent in the initiating posts, as compared to independent posts. These communicative verbs suggest explicit declaration of one’s communicative action for stance-taking, intertextuality, reformulation, or request of others’ communicative action for information seeking/sharing.

⁷³ <https://www.futurelearn.com/courses/moons/1/comments/516920>

⁷⁴ <https://www.futurelearn.com/courses/soils/1/comments/6450168>

6.3.9 Meta-language on learning and discussion

Meta-language keywords found in the initiating posts and independent posts are used to describe language or discourse referring to learning and discussion in the context of FutureLearn MOOCs.

Question and *article* are found to be used significantly more often in the initiating posts, *information*, *knowledge* and *course* are found to be used significantly more often in the independent posts. The usage of *question* and *article* in the initiating posts is elaborated on in relation to information seeking and sharing in the initiating posts.

6.3.9.1 *question*: Signposting information seeking

Of the 1281 occurrences of *question* in the initiating posts, most are used as nouns to refer to queries in information seeking or issues in stance expression. Using *question* to signpost their query may be a way to attract others' attention, especially when it is used at the start of a post in the pattern *I have a/question* (n=92, e.g., "I have a *question*. If there were 2 moons for the Earth, would the life be different than what it is today?"⁷⁵), or followed by colon (n= 104, e.g., "...*Question*: If climate change is making dung beetles go higher to live, does that mean there will be more dung left at lower levels?"⁷⁶). Sometimes, users seem to be self-deprecating by using negative adjectives including *stupid* (n=13), *silly* (n=11) and *dumb* (n=3) to introduce the question (e.g., "...This is possibly a *stupid question* but I couldn't find how to log out?..."), similar to the admission of one' being *wrong* or *missing* something, as discussed earlier.

Beside explicitly introducing a question, users also raise their opinions by using *question of* (n=74), *question about* (n=48) to focus on an issue (e.g., "...Sadly many of these smaller companies may push any accessibility improvements to one side in favour of spending a budget on, what they perceive as, more direct or profitable channels. It's a *question* of how to make companies realise the importance of accessibility I suppose..."⁷⁷).

⁷⁵ <https://www.futurelearn.com/courses/moons/1/comments/509432>

⁷⁶ <https://www.futurelearn.com/courses/soils/1/comments/6382940>

⁷⁷ <https://www.futurelearn.com/courses/digital-accessibility/2/comments/20266683>

6.3.9.2 *article and question*: Intertextuality and information sharing

Users mainly use *article* to refer to the readings provided in the course steps (e.g., “That was quite hard for me to comprehend the *article*...”⁷⁸). Similarly, *question* is also sometimes used to refer to a specific question in the quizzes or discussion prompts (e.g., “...I'm still floundering here and the half life *question* in the Week 5 assessment...”⁷⁹). This reference may help others to understand what the user is talking about, thus a common ground is created. Besides, *article* is also used to introduce what users have read on other websites when they share information (e.g., “This is a very useful *article*.www.care2.com/causes/the-solution-under-our-feet-how-regenerative-organic-agriculture-can-save-the-planet.html”⁸⁰). This practice of information sharing is further explored as URL-posting in Chapter 9.

6.3.9.3 Concluding remarks on meta-language

The meta-language keyword *question* in the initiating posts may function similarly to communicative verbs *tell* and *explain* discussed in section 6.3.7 when they are used to explicitly introduce users' questions. *Article* is also used to introduce information to be shared or sources of knowledge, similar to the function of *called* and hear-say evidentiality introduced by *told* and *says* discussed in section 6.3.7. In contrast, the meta-language keywords in the independent posts are typically used in conjunction with the independent keywords carrying positive evaluation, emotion and attitude to express appreciation (e.g., *useful/interesting information; enjoyed course*), or interest to learn (e.g., *expand/gain/broaden/improve/increase knowledge; hope/joined/look(ing) forward/taking course*).

It is worth noticing that meta-linguistic keywords found in both initiating posts and independent posts seem to be specific to this online space as they are all related to learning materials. This is unlike those meta-pragmatic expression discussed by Tanskanen (2007) that are on

⁷⁸ <https://www.futurelearn.com/courses/corpus-linguistics/1/comments/248012>

⁷⁹ <https://www.futurelearn.com/courses/moons/1/comments/678791>

⁸⁰ <https://www.futurelearn.com/courses/soils/1/comments/6361515>

users' commenting act. Nonetheless, as we shall see later, there are times that users use initiating keyword *sorry* to comment on their own posts.

6.3.10 Indefinite pronouns: *anyone* and *anybody* to call on others

Indefinite pronouns are used to refer to persons that language users cannot or do not want to specify exactly, and intend them to be addressed in a general and open way (Biber et al., 1999; Carter & McCarthy, 2006). *Anyone* and *anybody* are used significantly more often in the initiating posts, whereas *everyone* is used significantly more often in the independent posts. Both sets of keywords are used to address other users in the online discussion. This contrasts with one-to-one text messaging where *you* is used more often (Tagg, 2012), or in the replies in the online discussions where users address a specific user in a thread (Beers-Fägersten, 2008; Hewings et al., 2009; Sotillo & Wang-Gempp, 2016). *You* is a reply keyword that will be explored in Chapter 8.

In the following, the usage of *anyone* (n=860) and *anybody* (n=142) in the initiating posts is explained in relation to information seeking and sharing when they are used to address other users, and in relation to stance expression when they are used to refer to people in general.

6.3.10.1 Information/help seeking

Users use *anyone* and *anybody* to form questions in their request for recommendations or help from any users. 59% instances of *anyone* and 67% instances of *anybody* were used in a sentence ended with a question mark. For example, ".....So, does *anybody* have a good suggestion for a text book on Anaconda, Python and Pandas?"⁸¹; "I was unaware that there was a difference between ADD and ADHD. Can *anyone* clarify this for me? ..." ⁸²

Sometimes, the information or help seeking is framed as seeking shared experience. In these scenarios, users seem to be asking for advice or reassurance by seeking others who do the same as them (e.g., "I've logged my food intake as always but dont count the calories as i find it tedious , i

⁸¹ <https://www.futurelearn.com/courses/learn-to-code/1/comments/8989012>

⁸² <https://www.futurelearn.com/courses/dyslexia/1/comments/4514589>

have been eating healthy most of the time...*Anybody* else not counting calories?"⁸³), have similar problems (e.g., "Did *anyone* else have problems with links to Ted Talk? I was redirected and all 4 links went to the same short talk not related to the headings above?"⁸⁴), or have similar observations (e.g., "Is it just me or has *anybody* noticed, when you watch video's on soil erosion it's all happening on land that has been used for MONOCULTURES....."⁸⁵). This search for similar experience seems to function as an indirect way of asking for help, as replying users do indeed provide information needed.

6.3.10.2 Information sharing

Users also use *anyone* or *anybody* to offer recommendations in their initiating posts, without being prompted to do so (e.g., "...I highly recommend it to *anybody* who is interested in how the Public Health and Food Standard Agencies make these decisions ..."⁸⁶; "*Anyone* heard of Hildegard von Bingen's contributions to plantlore? That is one of the names to know when discussing the herstory of science...."). The qualification of *anyone* and *anybody* to those who are interested and those who have seen or heard about the recommendation reduces the force of imposition on others. This may also have a focusing effect on the issue to be raised, thus setting up a common ground.

6.3.10.3 Stance expression

Similar to information seeking, sometimes users frame their stance as if seeking people who share the same experiences (e.g., "*Anyone* ever had HAD to use the old fashioned earth toilet? I did when on holiday at the small holding of my Grandfather...."⁸⁷). Nonetheless, the common way of taking a stance by using *anyone* or *anybody* is by implying a meaning of "entirety", either in a rhetorical question (e.g., "How can it be that *anybody* can have children irrespective of their financial situation

⁸³ <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/19101897>

⁸⁴ <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/19363324>

⁸⁵ <https://www.futurelearn.com/courses/soils/1/comments/6386280>

⁸⁶ <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/19013150>

⁸⁷ <https://www.futurelearn.com/courses/ancient-health/1/comments/20010132>

/circumstances and expect society to pick up the bill?...’’⁸⁸) or in a statement (e.g., “*Anyone* that provides a product or service to another person should, at some level, be doing supply chain management...’’⁸⁹). In these occasions, *anyone* or *anybody* are used to refer to a general or unspecified person rather than addressing other users.

6.3.10.4 Concluding remarks for indefinite pronouns

Anyone and *anybody* are used in the initiating posts to engage and address other users, in both information seeking and sharing. Although this is less personal than using specific names to address others, it creates a dialogic relationship with other users since any user is invited to reply. Furthermore, the strategy of seeking shared experience also creates a dialogic relationship with those who have similar experience. Addressing other users, instead of facilitators, also suggests that users seemed to be aware of the social learning function of FutureLearn, i.e., learning via conversation with each other, instead of an educator-centred transmission model (R. Ferguson & Sharples, 2014).

The independent keyword, *everyone*, is also used to address other users. However, it is mainly used when users introduce themselves, as shown by the collocates *hello, hi, I am/'m, name*. It is possible that there are too many self-introductions at the start of the MOOCs where users are prompted to do so. The massive number renders this kind of self-introductory post less likely to receive replies than those establishing interpersonal relationships with *anyone* and *anybody* in their seeking of information or shared experience.

6.3.11 Polite speech-act formulae

Polite speech-act formulae, according to Biber et al. (1999), are those formulaic expressions used “in conventional speech acts, such as thanking, apologizing, requesting, and congratulating.” (p. 1093). They are typically used in conversation to signal an interaction among speakers, hearers and

⁸⁸ <https://www.futurelearn.com/courses/inequalities-in-personal-finance/1/comments/4439313>

⁸⁹ <https://www.futurelearn.com/courses/contract-management/4/comments/18392707>

messages, rather than contributing propositional meaning. Among the initiating keywords, *please* (n=589) and *sorry* (n=378) are such formulaic expressions.

6.3.11.1 *please* and *sorry*: Politely seeking information/help

Both keywords indicate politeness in information or help seeking (e.g., “*Please* can anybody help me on how to download the video on this page...”⁹⁰), although *sorry* may have a hedging function by apologizing for one’s own inability (e.g., “*Sorry*, but I don’t understand ex 6 the graphic representation of words. Can someone explain this to me please?”⁹¹). This is similar to the admission of one’s mistake discussed earlier. In-depth analysis of *sorry* also reveals its function in stance-taking and meta-pragmatic expressions, as elaborated on below.

6.3.11.2 *sorry* but I don’t agree: Sticking to one’s stance

sorry is used to preface or postface users’ strong stance or disagreement towards course content (e.g., “I’m *sorry* but I don’t agree with this. Maybe I’m misunderstanding the message but ‘more investment in health and education’, I assume she is talking about private investment...”⁹²). This apology can be seen as a polite behavior which shows concerns towards the “face” of the course designers who prepare the content (Brown & Levinson, 1987). It is also possible that this *sorry* is to anticipate any repercussion from others by apologizing in advance for one’s offensiveness in the initiating post. In this example, these interpretations can be further evinced by the user saying they “maybe ... misunderstanding the message”.

6.3.11.3 *sorry* in meta-pragmatic expressions

sorry is used in meta-pragmatic expressions that comment on one’s own posting or participation (e.g., “...*Sorry* to rant a little, but I’m sure we are not unique in this experience...”⁹³). This apology

⁹⁰ <https://www.futurelearn.com/courses/learn-to-code/1/comments/8572423>

⁹¹ <https://www.futurelearn.com/courses/dyslexia/1/comments/5144152>

⁹² <https://www.futurelearn.com/courses/inequalities-in-personal-finance/1/comments/4460408>

⁹³ <https://www.futurelearn.com/courses/palliative/1/comments/16952687>

regarding the nature of one's own postings could appeal to one's own "face" needs, such that others would not criticize the user but show support. At the same time, apologizing for personal emotion in one's posts suggests that users might perceive emotional expression as not allowed or off-topic in the learning setting, consistent with Laflen and Fiorenza's (2012) findings in an online distance learning course. In a somewhat similar yet different situation, users seem to apologize for being late in participating in the discussion (e.g., "*Sorry I am a little bit behind. I am just wondering under which circumstances will one type of law supersedes another?*"⁹⁴), as if it is not appropriate to be late, despite the fact that MOOC learning can be largely self-paced. These meta-pragmatic expressions indicate communication norms presumed by users in online discussions regarding emotional expression, long-windedness and late participation, which are the focus of their apology (Tanskanen, 2007). Nonetheless, these apologies could also be an expression of politeness to establish a dialogic relationship with others, assuming that the apologies are directed to an audience.

6.3.11.4 Concluding remarks on polite speech-act formulae

Please and *sorry* are often used in conversational speech acts (Biber et al., 1999), and may thus create a dialogic relationship with others, similar to the indefinite pronouns that are used to address others discussed earlier. Furthermore, instead of replying to a specific user, *sorry* is used in the initiating posts to preface users' disagreeing stances or potentially off-topic focus in the online discussions. This suggests that the users try to establish a dialogue with other users who will potentially read their posts.

In contrast, *thanks* and *thank* which indicate appreciation is found to be used significantly more often in the independent posts than in the initiating posts. They are often used with exclamation mark and another independent keyword *course* when users express appreciation to the

⁹⁴ <https://www.futurelearn.com/courses/contract-management/4/comments/18104444>

course providers (e.g., “...*thanks* to this course. It was a real eye opener...”⁹⁵). *yes*, which is often followed by *think* and *agree*, is another independent keyword that is often used in conversational speech act, and may be used to show agreement. The appreciation and agreement expressed towards course content can be an indication of user-content interaction, and is perhaps unsurprisingly found more often in the independent posts. The agreement practice in the independent posts will be discussed in Chapter 7.

6.3.12 Connectors

Connectors are used to link similar elements in a sentence. In the initiating posts, *if*, *or*, and *then* are found to be used significantly more often, whereas in the independent posts, *also* and *and* are used significantly more often. Four of these keywords, *or*, *then*, *also*, and *and* are normally used to coordinate similar ideas, in contrast to *if* which is used to elaborate on ideas (Biber et al., 1999; Carter & McCarthy, 2006). Specific attention is thus given to the keyword *if* (n=7404) in the initiating posts, especially *if*-conditionals that have been established as an important linguistic feature in argumentation and reasoning (Horsella & Sindermann, 1992; Louwerse, Crossley, & Jeuniaux, 2008), consensus and interpersonal relationship building in academic writing (Carter-Thomas & Rowley-Jolivet, 2008; Castelo & Monaco, 2013; Warchał, 2010), and politeness in spoken discourse (G. Ferguson, 2001; Moore, 2013).

6.3.12.1 *if you*: Establishing interpersonal relationship

If you (n= 887) is the most frequent bigram of *if*. Similar to *anyone* and *anybody*, *you* in the conditional refers to or addresses any potential user(s), and establishes interpersonal relationships with others. Users can use *if you* when sharing information to specify those who are interested or relevant (e.g., “*If you* visit the island of Kos here in Greece you can visit the site of the huge temple of Asclepius...”⁹⁶; “...*If you* are interested in Parliamentary language, you might also want to look at

⁹⁵ <https://www.futurelearn.com/courses/digital-accessibility/2/steps/155562/>

⁹⁶ <https://www.futurelearn.com/courses/ancient-health/1/comments/19586933>

an alternative site..."⁹⁷), thus creating a dialogic relationship with them. This specification also attenuates the imposition of their posts to others who are not interested.

If you, as well as *if I*, is also used to establish interpersonal relationships before users voice their stance, especially when the stance counters what is assumed in the current communicative context (e.g., "...*If you* look carefully you can see that it is not circular, it looks to me like it may be a polygon..."⁹⁸; "...*If I* may say so, we had a feudal system for far longer than..."⁹⁹). Besides countering, *if you* is used for focusing (e.g., "...*if you* read between the lines of Galens self aggrandisement this is the application of science and philosophy..."¹⁰⁰). In short, *if* is used to establish a dialogic relationship with others, before information or propositions are shared. This usage of *if you* might be similar to what Beers-Fägersten (2008) describes as one way of "seeking contact" (p. 222) by users in the hip hop interactive websites.

6.3.12.2 *if*-conditionals: Stance expression

Unlike the forms of *if you* and *if I*, *if-conditionals* are typically used in stance expression where *if* is used to specify conditions under which their argument stands (e.g., "...*If* the problem is not treated adequately it can result in low self-esteem..."¹⁰¹). Sometimes, the rhetorical force of the stance expression is enforced when *we*, *everyone*, *people* are used to generalize (e.g., "...Sustainability for all population is the target, this can be reached *if* everyone sees the sense in it..."¹⁰²; "...there is also an unfairness *if* people with any kind of disability have to be dependent on others..."¹⁰³; "...*If* we are so desperate for cash that we are willing to exploit such a special habitat I wonder what sort of world my children and grandchildren will inherit..."¹⁰⁴). Sometimes, *if you* is also used for

⁹⁷ <https://www.futurelearn.com/courses/corpus-linguistics/1/comments/339785>

⁹⁸ <https://www.futurelearn.com/courses/moons/1/comments/567400>

⁹⁹ <https://www.futurelearn.com/courses/inequalities-in-personal-finance/1/comments/4420764>

¹⁰⁰ <https://www.futurelearn.com/courses/ancient-health/1/comments/19606238>

¹⁰¹ <https://www.futurelearn.com/courses/dyslexia/1/comments/4630356>

¹⁰² <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/19189514>

¹⁰³ <https://www.futurelearn.com/courses/digital-accessibility/2/comments/19464383>

¹⁰⁴ <https://www.futurelearn.com/courses/exploring-our-oceans/1/comments/483000>

generalization when *you* is used to refer to people in general (e.g., “...*If you* have a great income, you can increase the wealth...”¹⁰⁵). Although these propositions sound like generalization, they can arguably be interpreted as a speculation or a suggestion to other users, given that the condition specified by *if* is one of the many alternatives.

6.3.12.3 Concluding remarks on connectors

Among the connectors found to be initiating keywords, *if* is analysed in detail for its usage in setting up hypothetical conditions, i.e., *if-conditionals*. When *if you* and *if I* are used, it is mainly a way of establishing interpersonal relationships with others while sharing information or stance. When not used with *you* and *I*, *if* is used to specify the condition under which a stance stands, implicitly acknowledging other possibilities, allowing alternative conditions to be voiced by other users.

6.3.13 Other functional groups of keywords

Besides the abovementioned functional groups, the communicative function and meaning of seven other groups of keywords are only briefly introduced here. They are reproduced in Table 6.4. They include boosters and quantifiers used for qualifying stance (e.g., “...*Surely* widespread campaigns based on the economic savings might influence behaviour change...”¹⁰⁶); activity verbs used for narrating events or actions (e.g., “...Many times I have *tried* to get on a Geography Course to study at some level...”¹⁰⁷); and pronouns, punctuation marks and grammatical words that have high frequency of occurrence in the corpus and are involved in a wide range of communicative functions which cannot be easily categorized. Lastly, a group of keywords remains uncategorized for two reasons. Firstly, they are the only keywords with a specific function, for example *why* is the only *wh*-question word among the initiating keywords. Secondly, some of the keywords carry multiple meanings and functions in the corpus, and no salient function can be concluded. For example, *well*,

¹⁰⁵ <https://www.futurelearn.com/courses/inequalities-in-personal-finance/1/comments/4283361>

¹⁰⁶ <https://www.futurelearn.com/courses/exploring-our-oceans/1/comments/495587>

¹⁰⁷ <https://www.futurelearn.com/courses/soils/1/comments/6417917>

an initiating keyword, can function as preposition in “*female as well as male*” and evaluative in “*feeling well*”; *just*, another initiating keyword, can mean “only” in “*I have just one question*”, “recently” in “*I have just read*”, “*simply*” in “*I could just sit at home*” etc. In the following section, I focus on two related initiating keywords that are uncategorized *example* and *e.g.*

Table 6.4 Seven functional groups not elaborated in the text

Functional grouping	Initiating keywords	Independent keywords
Activity verbs	used, tried, came	joined, affects, helps, achieve, work, gain, meet, improve
Quantifier	any	all, lot, much, every
Booster	surely, rather, else	really, very, definitely, always
Pronouns	he	I, my, our, their
Grammatical	the, that, there, here, does, did, was, were, 's, on, by, than, same	am, 'm, have, for, about, with, to, more
Punctuation	,...-();?":	!.
Uncategorized	example, e.g., 1, one, two ¹ , numbers, why, whether, just	like, well, week, main, currently, working, opportunity, education, environment, mind

¹1, one, and two arguably function as quantifiers as well, but they differ from the other quantifiers in the sense that they are numerals that specify exact amount (Biber et al, 1999) and do not have the intensifying or down-toning function in stance expression.

6.3.14 Uncategorized keywords: *example* and *e.g.*

example and *e.g.* have been established by previous research (Biber, 2006) as a means to introduce exemplification in informational presentation in spoken and written language. However, *example* (n=1323) and *e.g.* (n=536) in the initiating posts are not categorized into any functional group because they are found to carry multiple functions in the initiating posts, including acting like connectors, parenthetical insertion and meta-language. These functions are explained in the following especially in the case of *for example* (n=828), which is the most frequent bigram with initiating keyword *example*.

6.3.14.1 *for example*: Linking to a typical case

for example is typically used to link claims or questions to exemplar cases or entities. When used in a bracket, it seems to be mentioning entities in passing, especially in questions (e.g., “Once you have loaded a set of files (the BROWN files, *for example*), and you now wish to analyse a different set, is there a way of?”¹⁰⁸). This mention in passing is typically not referred to again in the replies to the initiating posts. This function is mainly for illustration purposes.

However, when *for example* is used to start a sentence, it often introduces an elaborated case for exemplification following the presentation of questions or stances. Unlike mentioning the example in passing, as shown above, users use *for example* to elaborate on their questions (e.g., “What is the best way to present data in a spreadsheet when there are different categories for the same data? *For example* I'm using suicide rates data and it has columns with headings ”¹⁰⁹).

When expressing stance, the cases following *for example* can be hypothetical or real experience, as shown below:

“.....I do have a strong belief that nurture can play a part both with regards to our own choices but also the influence of others on our actions and behaviours [*stance*]¹¹⁰. *For example*, if you grow up in a home that encourages trying different foods and eating healthily, you will be likely to incorporate this into your own lifestyle as you become more independent [*hypothetical example*]¹¹¹.....”¹¹²

“As part of a team at a UK university in promoting online learning systems for assessment, I think businesses should focus much of their energy into ensuring their customers and staff want to use technologies [*stance*]¹¹³. *For example*, we have experienced first-hand that if

¹⁰⁸ <https://www.futurelearn.com/courses/corpus-linguistics/1/comments/262907>

¹⁰⁹ <https://www.futurelearn.com/courses/learn-to-code/1/comments/8711289>

¹¹⁰ My emphasis to point out the part of the initiating post where stance is expressed.

¹¹¹ My emphasis to point out the part of the initiating post where example is elaborated.

¹¹² <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/19450586>

¹¹³ My emphasis to point out the part of the initiating post where stance is expressed.

lecturers are enthusiastic and engaged with our online learning environment then our students are more inclined to be in favour of using technology.....[real example]¹¹⁴¹¹⁵

6.3.14.2 *for example*: Parenthetical insertion

When *for example* is used in the middle or at the end of a sentence, it seems to be an optional or parenthetical element and may hedge what has been said (see Figure 6.6 for concordance lines).

This could be a signal that it is just one of the situations or examples, unlike the elaborated example for supporting stance, as introduced by *for example* at the beginning of a sentence. It could also be a politeness insertion suggestive of the interactive nature of the initiating posts.

Figure 6.6 Concordance lines of “*for example*” used as parenthetical expression.

stop as a result of climate change, due (*for example*) to increased flows of fresh water into need both hands to propel my wheelchair, so (*for example*) on the Underground, I find it difficult here, as I always thought that with beginners (*for example*) you would want to ask more basic ties to people that have motor impairment. Paris, *for example*, is a (bad) example: most underground stations weather- and grain supply ships from Alexandria, *for example*, did not sail during the winter months- Is there a 'Rename()' function that allows you, *for example*, to rename the column heading "SP.POP. full advantage. While a concordance search: *_NN1, *for example*, will generate all those singular nouns, the much more, by providing lists of new terms, *for example*, and allowing certain students more time to the risks that would be calamitous. One might, *for example*, consider insurance with lower premiums but high effort on the result, reduced corruption, Singapore, *for example*, for 50 years become one of the first the label "disease", although if the cause is, *for example*, compulsive eating, that is considered a disorder throughout the heavy bombardments? We are not saying, *for example*, that the impactors were larger during the epas zevinæ Newman, 1979) looks like a plant as, *for example*, do fossil crinoids. Is there a known

6.3.14.3 Adjective + *example*: Meta-language on learning and discussion

The second most frequent bigram of *example* is determiner (*an, the, one, this, another*) + *example* with 246 occurrences, followed by *example of* with 180 instances. Another pattern worth noting is determiner + adjective + *example* with 86 occurrences, of which the adjectives are mainly positive

¹¹⁴ My emphasis to point out the part of the initiating post where example is elaborated.

¹¹⁵ <https://www.futurelearn.com/courses/digital-accessibility/2/comments/19418858>

(e.g., *good, great, classic*). The present analysis focuses on *adjective + example* to illustrate the meta-linguistic function of *example*, similar to the keyword *question* which is used to signpost a subsequent question. Most of the time, the example is given in relation to the course content or discussion prompt and to support a stance. Sometimes, the example itself becomes a topic of discussion in the thread initiated, as shown in Figure 6.7.

Figure 6.7 Thread 4442056

Thread 4442056

Source: <https://www.futurelearn.com/courses/inequalities-in-personal-finance/1/comments/4442056>

Initiating Post 2015-04-12 18:21:56 Like: 4

User f1-318

How refreshing to see a government providing affordable homes for their people!! Meanwhile here in the UK, councils are selling estates to private firms who are evicting the tenants (mostly there through social housing) bulldozing the homes and 're building luxury houses in its place. (A *fine example* can be found in Barnett, the Sweets Way estate).

Reply 1 2015-04-13 14:53:35 Like: 0

User f1-58

Yes, apparently the tenants on Sweets Way estate are expecting to be evicted at some time unknown to them this week. Are we back in the 17th century? Not a very sweet way to be treated. There has been a petition against this and they are asking for support from anyone living nearby.

Reply 2 2015-04-13 15:32:47 Like: 3

User f1-318

I live nowhere near Barnett, and I know no one on the estate, but I have been helping where I can, supporting their petition and spreading the word. Its absolutely barbaric!

Reply 3 2015-04-24 10:44:27 Like: 0

User f1-18

But we saw that the level of social housing is still relatively high in the UK.

In Figure 6.7, a *fine example* is used by the initiator to introduce a case of current affairs (“the Sweets Way”) to align with their stance that “in the UK, councils are selling estates...”. This example is also raised to contrast with the case study presented in the course content about public housing in Singapore (not shown here). This example is not elaborated on in the initiating post, but is expanded on by another user in Reply 1, who also aligns with the initiator, “Yes, apparently”. Nonetheless, another user voices an opposing view by saying “the level of social housing is still relatively high”.

Besides current affairs in Figure 6.7, the example can also be a historic event, as shown in Figure 6.8.

Figure 6.8 Thread 19531657

Thread 19531657

Source: <https://www.futurelearn.com/courses/ancient-health/1/comments/19531657>

Initiating Post 2017-02-09 10:01:22 Like: 3

User ah1-644

Competition among doctors and the various suggested remedies they advanced have never really left us although better knowledge and diagnosis has reduced the scope for this. *A very good example* of competition was in the treatment of George 111's "madness" in 1788. No-one really understood the cause- thought now to be porphyria- and the remedies were varied

Reply 1 2017-02-09 16:35:20 Like: 0

User ah1-973

I recently read the satirical 19th century novel "Doctor Thorne" by Anthony Trollope in which a major theme is the rivalry for wealthy patients between doctors - several of which are more concerned with their public reputations as the best than for the health or their patients.

In Figure 6.8, the example given is a historic event “George 111’s “madness””, which is in turn supplemented by another user with another example which is a novel, “Doctor Thorne”. The historic event is introduced by “A very good *example*” after the initiator takes a stance regarding “Competition among doctors.....”. The introduction of examples in Figure 6.7 and Figure 6.8 can also be considered as information sharing, although no exact source is provided, unlike the URL-posting to be explored in Chapter 9.

Besides real examples, as in the case of current affairs in Figure 6.7 and the historic event in Figure 6.8, the example shared can also be personal experience, as shown in Figure 6.9.

Figure 6.9 Thread 19347934

Thread 19347934

Source: <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/19347934>

Initiating Post 2017-02-02 23:18:49 Like: 11

User n4-2501

You are no more limited by your epigenetic markers than you are by your DNA structure. Epigenetics is the impact of food and other environmental factors on gene expression and gene suppression. You can have a deadly mutated gene, but if it isn't expressed it won't cause any harm. Sounds simple, but much more complex in real life. I look at the impact of epigenetics in my family as *a good example*. My immediate and extended family has a predisposition to obesity and is populated with a high percentage of very obese and Type 2 diabetics. I've been able to avoid this by regular exercise and a good diet, while my brothers carried a lot of excess weight and suffer from the chronic problems associated with diabetes.

Reply 1 2017-02-04 20:20:44 Like: 2

User n4-3302

Great observation and well done on taking control of your health journey.

I'm on a similar path, taking responsibly for my health, and it is encouraging to know that we are able to halt or reverse some of the health problems that through epigenetics may have been passed to us from our parents and grandparents.

In Figure 6.9, the example given is a detailed personal experience “epigenetics in my family” to support the initiator’s stance, which is then complemented and complimented by a replying user who is “on a similar path” - “Great observation and well done”. This and other cases show that users share examples in their initiating posts in relation to what they learn from the courses and their stance towards the content, indicative of user-content interaction as well as information sharing practices. The examples in the initiating posts seem to encourage others to share their examples in their replies, as shown in Figure 6.8 and Figure 6.9, similar to the practices of telling second stories in conversations for stance-taking (Siromaa, 2012) or experience talk in online discussions in e-learning (Kääntä & Lehtinen, 2016), whereas the example in Figure 6.7 forms the topic for other users to discuss. In all these instances, the examples specify one of the many conditions that are in line with

the stance stated, as with the *if*-conditionals explained in section 6.3.12, while expanding the dialogic space for other examples to be pitched in. Sometimes, it appears that users can also be primed by discussion prompts containing the word *example* to use the word in their posts. These discussion prompts typically ask users to provide examples based on their personal experience or knowledge (e.g., “If you have to work with data, you may have stories about dirty data you have seen or been presented with. If you’re at liberty to do so, share some of those *examples* in the comments.”¹¹⁶).

Based on the detailed analysis on *for example* and *adjective + example*, it can be concluded that these keywords have more than one function in the initiating posts, including linking function like connectors, metalinguistic reference to course content and parenthetical insertion. Thus, they are grouped as uncategorized. It should therefore be recognized that other uncategorized keywords not analysed here may also have multiple functions which can be further dissected.

6.4 Discussion

In this section, general patterns in the discourse of initiating posts and independent posts based on the keyword analysis are first summarized to reveal their differences, thus addressing the RQ1 regarding the difference in linguistic features and discourse practices between these two types of posts. These general patterns are also compared to the existing literature that focus on linguistic features in new posts that predict the chance of receiving replies. Although the finding is compared to these previous studies, this thesis takes the view that linguistic features need to be examined for how they are used in discourse practices, as explained in Chapter 4. Therefore, after the comparison, the dialogic nature of the discourse practices in the initiating posts are discussed to explain how a conversation is established in online discussions, while discourse practices of independent posts will be further explored in the next chapter to understand their functions in the online discussions. The

¹¹⁶ <https://www.futurelearn.com/courses/learn-to-code/1/steps/42103>

linguistic features of independent posts are first discussed because of the conflicting findings that they are found in initiating posts in previous studies.

6.4.1 General patterns of independent posts

Compared to initiating posts, a significant number of the independent posts include self-references when users introduce themselves, talk about their intention to join the MOOCs, and express appreciations and reflections regarding their learning, as realized by mental verbs carrying a positive emotional meaning (e.g., *looking forward, enjoy, enjoyed, love*) or relating to learning (e.g., *understanding, learned*), and positive evaluative adjectives (e.g., *interesting, informative, great*), collocating with the first person pronouns (e.g., *I, we*). The self-references are typically prompted by the discussion prompts at the start of the MOOC, where users are encouraged to make their first post in the online discussions to introduce themselves, as realized in their addressing others with *everyone*, and towards the end of the MOOC where they are encouraged to reflect on their learning, following the design of FutureLearn (R. Ferguson & Sharples, 2014). Although the analysis was arguably skewed by these self-references triggered by the learning design, it is not possible to select only the comments that are independent of the discussion prompts and I can risk cherry-picking. Therefore the whole corpus of users' comments is analyzed to ensure the data-driven approach taken in this thesis reflects what happens in the online discussions.

These self-references are found significantly more often in independent posts, that do not receive replies, in contrast to previous findings on Usenet groups or news websites where autobiographical references are more likely to generate replies (Arguello et al., 2006; Burke et al., 2007; Ziegele et al., 2014). This difference may be because most of the self-references in Futurelearn MOOCs discussions are concentrated at the beginning of the courses where there are an overwhelming number of posts, such that the chance of getting a reply is reduced (Himmelboim, 2008). It is also possible that the MOOC discussion space is unlike Usenet groups which are generally sustained by users who have long been committed to the group, and therefore value and reply to

other users' self-references, especially those made relevant to the history of the community (Burke et al., 2007). In contrast, the online discussion in a FutureLearn MOOC does not have a long history but is limited to the period when the MOOC is running and users often do not have recurrent interactions with the same users over the course, therefore it is hardly possible to make a self-reference in relation to the history of the community (Sunar et al., 2015).

However, personal experiences as introduced by *example*, the initiating keyword, suggest that self-references in relation to a specific content might still trigger replies from others, similar to what Arguello et al. (2006) call testimonials. Similarly, in cancer support group, first person pronouns are found to predict the chance of receiving replies, possibly because they are related to personal experience related to the cancer (Crook et al., 2016). This is unlike the self-reference in the independent posts in the MOOC discussion space that seem to be in response to the prompts at the start of the course ("Before you move on, why not introduce yourself now in the comments?"¹¹⁷), which could be generic rather than in relation to a specific issue.

The fact that appreciation and reflection are also commonly found in the independent posts towards the end of FutureLearn courses is in contrast to Arguello et al.'s (2006) finding that expressions of positive emotion increase the chance of receiving replies. These expressions may not receive replies due to their overwhelming number (Himmelboim, 2008). Most importantly, the discourse of the posts may not be designed to establish dialogue with other users, i.e., user-user interactions, but serves to engage in user-content interactions as they are typically in response to the prompts. These different findings suggest the importance of examining the discourse practices which employ particular linguistic features, rather than just stop at the linguistic features as in previous studies (Arguello et al., 2006; Crook et al., 2016). The discourse practices of independent posts will be further explored in Chapter 7.

¹¹⁷ <https://www.futurelearn.com/courses/moons/1/steps/3793>

6.4.2 General patterns of initiating posts

The keyword analysis shows three general patterns in initiating posts; that is, information seeking and sharing, meta-linguistic expressions and stance-taking. I focus here on information seeking and stance-taking in order to compare my findings to the existing literature, because information sharing practices and meta-linguistic expressions have not been discussed in previous studies.

The top initiating keywords based on effect size are those relevant to seeking information or help, consistent with previous findings that questions have a higher chance of receiving replies (Arguello et al., 2006; Chen et al., 2020; Rooderkerk & Pauwels, 2016; Ziegele et al., 2014). However, unlike previous studies that see asking questions as a homogenous practice, the functional grouping and discourse analysis of the initiating keywords in this thesis reveals that a number of discourse practices are used in seeking information or help, instead of just bald requests. The discourse practices found include making concession to one's mistake, displaying "unknowing" status or "interest to know", seeking similar experience with the use of indefinite pronouns, and elaborating with examples. This reflects the findings by Burke and colleagues (2007; 2008) who observe that requests with reference to online community history and politeness strategies are more likely to receive replies, although the discourse practices found in this thesis differ. This again may be because of the difference between Usenet groups and MOOCs. Usenet groups are made up of members who might value their relationships with other members, whereas MOOCs may be more oriented to epistemic expression regarding users' "knowing" status, given its learning orientation (Burke et al., 2007; Gillani & Eynon, 2014; Sunar et al., 2015).

The other general pattern found are practices relevant to stance-taking, which may indicate the start of stance (dis)alignment in the user-user interactions prompted by the initiating post (Du Bois, 2007). These general patterns are realized by various linguistic features in the initiating posts, including modals, hedges, expression of "partially knowing", "unknowing", negation, indefinite pronouns and *if*-conditionals. Both Arguello et al. (2006) and Crook et al. (2016) find a category called "cognitive mechanism" that predict the likelihood of receiving replies. Within the category,

Crook et al. (2016) reveal subcategories of tentativeness and certainty. It is possible that the tentativeness category might contain modals and hedges as found in the current study. However, as mentioned earlier, the earlier studies do not reveal the word forms found or explain the discourse practices in these categories. Therefore, there is no way to further investigate the difference in findings between the current study and previous studies.

Lastly, both Arguello et al. (2006) and Crook et al. (2016) find that negative emotion expressions also increase the chance of receiving replies, whereas the keyword analysis in this study does not reveal any keywords related to negative emotions, except the negation and probably *wrong* and *missing* which seem to be used more often for concession or expressing “unknowing” status in this online space. Negation in this study, especially for stance-taking, may be similar to the disagreement found in Chen and Chiu (2008) and Chen et al.'s (2020) studies that find that disagreement with a previous message is more likely to trigger more messages. However, these studies do not reveal how the disagreement is realized. In contrast, the current study reveals that users express opposing stance towards course contents in their initiating posts while utilizing mitigating strategies to maintain relationship and leave room for conversations, similar to what Drasovean and Tagg (2015) found in TED commenting space. How these linguistic features realize different discourse practices for stance-taking will be discussed in the next section. In doing so, the current study extends previous findings (Arguello et al., 2006; Chen & Chiu, 2008; Chen et al., 2020; Crook et al., 2016) by investigating further the linguistic features in context to understand the discourse practices that are likely to generate replies, thus exploring the dialogic nature of online discourse while addressing RQ2 regarding the discourse practices that can potentially initiate dialogic conversations.

6.4.3 Discourse practices in initiating posts that are of dialogic nature

6.4.3.1 Indicating one's epistemic status: intention to know, “partially knowing”, “unknowing”

As shown by the analysis of the initiating keywords, users indicate varying degrees of epistemic

status to establish a dialogic relationship with those who have a “knowing” status, either for seeking

information or expressing their stance (Heritage, 2012). This practice is realized by explicit expressions of their intention to know (e.g., “I would like...”, “I would be interested”, “I wonder”), or admission of their “unknowing” status or incapability (e.g., “sorry”, “I might be wrong”, “missing something”, “I don’t know/understand”, “I could not”), whereas a “partially knowing” status is typically expressed by modals or hedging, either by self-attribution to indicate one’s subjectivity (e.g., “I would say”, “it seems to me”) or distancing oneself (e.g., “perhaps”, “it would/might/could”, “it seems”).

In information seeking, the intention to know or “unknowing” status may create a dialogic space for others to fill with what they know or with expression of their similar “unknowing” epistemic status. In stance-taking, the expressions of “unknowing” or “being wrong” may be one way of making concessions to avoid voicing an absolute certainty that could stifle other voices (Concannon et al., 2017; Grant, 2010), and could also be one way of introducing an unexpected or dispreferred stance to the current communicative context in a way that reduces one’s own “face” threat when others do not agree (Baumgarten & House, 2010). On the other hand, the “partially knowing” status can indicate uncertainty which creates a dialogic space that recognizes other alternatives while voicing one’s own stance (Martin & White, 2005). Either way, the resulting dialogic space may thus encourage replies from other users, thus starting a user-user interaction.

6.4.3.2 Addressing potential readers to realise dialogic nature of online discourse

As reviewed in Chapter 3, one aspect of the dialogic nature of human language (Bakhtin, 1981) is that language users design their utterances as if they are addressed to someone, real or imaginary. In the initiating posts, several keywords seem to be overtly used for this practice. Politeness formula including *please*, indefinite pronouns *anyone* and *anybody*, and *if you* are used to create a dialogic relationship with other users who potentially read the posts, despite the fact that users cannot specify names among an unknown group of users at the start of the threads. This is sometimes compensated for by stating who might be relevant, for example “*If you are interested in*”. This

strategy has been used for both information seeking and sharing. These keywords, as well as the metalinguistic use of *question*, also have a focusing effect by highlighting what a user wants the other users to pay attention to.

Furthermore, the politeness formula *sorry*, parenthetical insertion *for example*, and *I mean* indicate users clarifying their own comments as if they were responding to an imaginary audience's request to clarify. Specifically, *sorry* appearing in meta-pragmatic expressions suggest users explicitly try to mitigate any potential problems they might create for others, such as long posts, indicative of the interactive nature of their posts (Tanskanen, 2007). *I mean* and *for example* are used to reformulate what has been mentioned before, either by the users themselves or others in the prior communicative context. This reformulation is similar to the involvement strategies in oral conversations that establish a dialogic relationship between previous and future utterances (Tannen, 2007).

6.4.3.3 Intertextuality

As reviewed in Chapter 3, intertextuality reveals the way in which language users relate to others' utterances or the wider socio-cultural context (Fairclough, 2003). In the initiating posts, keywords indicating hear-say evidentiality, *says*, *told*, and *called*, and meta-language *article* and *example* are used to introduce sources or experiences as explicit indicators of intertextuality for both information sharing and stance-taking. This intertextuality can supplement the user's own voice with additional voices from others by attribution to third parties, thus expanding the dialogic space. This is consistent with Himmelboim et al's (2009) findings that posts importing content from elsewhere on the internet trigger user-user interactions, but inconsistent with Rooderkerk and Pauwels (2016) who found hyperlinks do not necessarily increase the chance of receiving replies. The practice of URL-posting will be explored further in Chapter 9.

6.4.3.4 Setting up a shared space with multiple voices

As argued in Chapter 3, the dialogic space also forms a shared space for intersubjectivity to develop.

In the initiating posts, *if*-conditionals are used to create hypothetical situations or counterfactuals to speculate on causes and consequences, whereas *example* and *for example* are typically used to present real life supporting evidence. Both provide a concrete situation as a common ground for others to comment on (G. Ferguson, 2001) or for an exchange to develop (Landqvist, 2016; Liu & Liu, 2017). Furthermore, given that *if* and *example* refer to one of many alternative situations, the hypothetical, counterfactual situation or real-life example may trigger others to put forward more situations or examples, thus expanding the dialogic space. Lastly, negative propositions embedded with *n't*, *cannot*, *can't* may also invoke the positive counterpart, thus setting up a dialogic space of multiple views. The negation may also introduce controversial and unexpected stances that tend to increase the chance of receiving replies (Rooderkerk & Pauwels, 2016; Ziegele et al., 2014).

6.5 Conclusion

Using a data-driven corpus linguistics approach, this chapter found significant lexical differences between initiating posts and independent posts, which in turn point to distinct discourse practices, thus addressing the RQ1 on the differences between these two types of posts. More importantly, the analysis extended previous studies that only investigate linguistic features by showing how they are used in context and in discourse practices. By doing so, it was argued that discourse practices may expand or contract the dialogic space, thus impacting on the potential of the post to receive a reply. This investigation of discourse practices thus addresses the RQ2 on how users' discourse practices initiate dialogic conversations.

The initiating keywords and functional grouping revealed the particular ways that users go about stance-taking, information-seeking and sharing in the online discussions. By creating an epistemic or dialogic relationship with potential audiences, users increase the chance that their posts can attract replies. Furthermore, by establishing a shared space or referring intertextually to

alternative sources, a dialogic space that welcomes other voices may be opened up. These discourse practices contrast with those found in the independent posts where users make more self-references for expressing appreciation and reflection. Although the discourse practices in the independent posts might not open up a dialogic space, thus not receiving replies, these discourse practices may be employed by users to engage in interaction with course-content instead of potential audiences. This possibility will be explored in the next chapter.

Lastly, although this thesis is mainly concerned with online discourse in general, it is also worth relating the current findings to the previous findings on MOOCs, the research setting for online discourse in this thesis. As reviewed earlier, previous MOOC studies typically categorize comments for *what* they are, for example on-topic (Cui et al., 2017) and question (Poquet et al., 2018). Assuming that stance-taking and information sharing are on-topic, and information seeking relates to question, the current findings show that users employ various discourse practices in these on-topic comments and questions, while creating a dialogic relationship with others. This suggests that a comment can contain elements both for on-topic discussion and for establishing social relationships, such that these two elements are not necessarily separated but can be construed at the same time. The approach taken in this chapter therefore extends previous MOOC research by pointing to the formal and functional complexity of users' posts and thus the need to drill down into individual posts in order to fully understand how users interact in MOOC discussion spaces.

Chapter 7

Independent Posts: Dialogic contraction and/or user-content interactions?

7.1 Introduction

This chapter explores discourse practices in the independent posts, based on the keyword analysis and functional grouping conducted in Chapter 6. The investigation of independent posts, which despite not receiving replies, will also indirectly address the second research question from an opposite scenario:

RQ2: How do these discourse practices initiate, sustain or hinder dialogic conversations in online discussions?

In contrast to initiating posts, independent posts may fail to attract others' replies despite their authors' intentions, or be written mainly in response to the course content, thus indicating user-content interactions (Ksiazek & Lessard, 2016; Ziegele et al., 2014). The functional grouping and discourse analysis of independent keywords may thus reveal discourse practices that might contract a dialogic space for others' voices but are designed to engage in a dialogue with the page content.

This chapter starts by describing the functional grouping of the independent keywords which shows the general patterns in the independent posts. An in-depth analysis of two keywords, *think* and *agree*, are then conducted because of their significance in stance-taking as established by previous research in other online spaces (e.g., Baym, 1996; Bolander, 2012; Myers, 2010). The analysis reveals how stance-taking practices in the independent posts differ from those in the initiating posts such that nobody responds to their stance to continue a discussion. This chapter

concludes by arguing that although the discourse practices found in the independent posts are less likely to start user-user interactions, they might reveal the discourse practices by which user-content interactions are realized. The user-content interactions can be seen as users' internal conversation while interacting with the content of the course according to the design of FutureLearn (R. Ferguson & Sharples, 2014), thus potentially successful engagement from the point of view of the host of the online space.

7.2 Functional grouping of independent keywords: Expressions of appreciation

The functional grouping of the independent keywords is reproduced in Table 7.1, with the initiating keywords as comparison. Several observations can be made about the independent keywords in contrast to initiating keywords. First, unlike initiating keywords, none of them are hedges, negative particles, or communicative verbs. Second, all but one evaluative independent keywords are positive (e.g., *easy*, *excellent*, *interesting*), compared to the negative evaluative keywords in the initiating posts. Third, the mental verbs found to be independent keywords are mainly positive emotional and attitude verbs (e.g., *enjoy*, *hope*, *looking forward*), compared to the two epistemic mental verbs (*wonder*, *wondering*) found in the initiating posts which are used to express "unknowing" status. Fourth, quantifiers indicating high intensity are among the independent keywords (e.g., *all*, *lot*, *much*). Lastly, three first person pronouns (*I*, *my*, *our*) are found to be independent keywords, suggesting self-references in independent posts.

Table 7.1 Functional grouping of independent keywords and initiating keywords

Functional grouping	Independent Keywords	Initiating Keywords
Modals/Modal expression	will, need, able ¹	might, would, could
Hedges		perhaps, seems, sort
Evaluative	easy, excellent, better, interesting, informative, great, important, good, new, difficult	wrong, missing
Negation		cannot, ca, n't
Mental verbs	aware, understanding, learned, think, agree, feel, feeling, keen, hope, hoping, looking, forward, enjoy, enjoyed, love	wonder, wondering
Communicative verbs		mean, explain, tell, says, say, told, called
Activity verbs	joined, affects, helps, achieve, work, gain, meet, improve	used, tried, came
Meta-language on learning and discussion	information, knowledge, course	question, article
Indefinite pronouns	everyone	anybody, anyone
Polite speech-act formulae	yes, thanks, thank	please, sorry
Connectors	also, and	If, or, then
Quantifier	all, lot, much, every	any
Booster	really, very, definitely, always	surely, just, rather, else
Pronouns	I, my, our, their	he
Grammatical	am, 'm, have, for, about, with, to, more	the, that, there, here, does, did, was, were, 's, on, by, than, same
Punctuation	!.	,...-();?'"':
Uncategorized	like, well, week, main, currently, working, opportunity, education, environment, mind	example, e.g., 1, one, two, numbers, why, whether

¹ 96% of the instances of *able* collocated with *to*, forming the modal expression *able to* (Carter & McCarthy, 2006).

In the independent posts, the keywords of positive evaluations, emotional and attitude verbs, quantifiers and boosters, meta-language, first person pronouns, politeness formulae and the exclamation mark (!) attest to the positive sentiments and appreciation expressed by users in the online discussion. Most of the positive sentiments constitute users talking about their learning. For example, what the users want to learn (e.g., “... I 'm *really looking forward* to learn more about computer-mediated communication...”¹¹⁸; “...I'm *really* interested in finding out more *information* concerning dyslexia...”¹¹⁹), or on what they have learnt (e.g., “...I'm *looking forward* to taking some of the principles I have *learned* here and applying them to...”¹²⁰; “Again, another *great week* with useful and relevant *information*.”¹²¹), as prompted respectively by the learning activities or contents at the start (e.g., “Describe your interest in corpus linguistics and if you wish, tell us what you hope to get from this course.”¹²²) and at the end of each course (e.g., “Acknowledgements - This course is the result of a collaborative process.”¹²³). Sometimes, the positive sentiments can be expressions of gratitude to the course educators (e.g., “*Excellent* range of resources, *thanks!*”¹²⁴). The fact that the appreciation is apparently directed towards the course facilitators and educators may be why other users do not reply.

Besides appreciation, these expressions also suggest that users engage with the course content, especially when users compliment the courses or specific materials, as shown in the examples above. Furthermore, the expressions of wanting to learn at the start of the course and the expressions of having learnt something at the end of the course attest to their “becoming knowing” epistemic status. The first-person pronouns *I*, *my*, *our*, along with the epistemic expressions,

¹¹⁸ <https://www.futurelearn.com/courses/corpus-linguistics/1/comments/240881>

¹¹⁹ <https://www.futurelearn.com/courses/dyslexia/1/comments/4510497>

¹²⁰ <https://www.futurelearn.com/courses/contract-management/4/comments/18390658>

¹²¹ <https://www.futurelearn.com/courses/dyslexia/1/comments/4807274>

¹²² <https://www.futurelearn.com/courses/corpus-linguistics/1/steps/5028>

¹²³ <https://www.futurelearn.com/courses/moons/1/steps/4856>.

¹²⁴ <https://www.futurelearn.com/courses/dyslexia/1/comments/4821957>

understanding, aware and learned are used to indicate such epistemic status (e.g., “I *enjoyed* this course and *definitely learned* a lot in...”¹²⁵).

The appreciation and the expressions of “becoming knowing” suggests the reflective nature of the independent posts, thus may not necessarily invite responses from others. “Becoming knowing” indicates an epistemic change rather than establishing any epistemic relationship with others. This contrasts with the initiating posts which consist of discourse practices indicative of “unknowing” and “partially knowing” status that open a dialogic space for others, as discussed in Chapter 6.

7.3 Analysis of selected keywords: *think* and *agree* for stance-taking

Besides the positive emotional and attitude verbs, *think* and *agree* are two other mental verbs found to be the independent keywords. *think* and *agree* are well-established as frequently used in stance-taking, either in spoken discourse (Kärkkäinen, 2003; Pomerantz, 1984) or internet-mediated communications (e.g., Baym, 1996; Bolander, 2012; Hewings et al., 2009; Myers, 2010; Sotillo & Wang-Gempp, 2016). These researchers have also argued for the dialogic nature of *think* and *agree*. However, contrary to this argument, *think* and *agree* are keywords of the independent posts that do not receive any replies, rather than keywords of the initiating posts that start user-user interactions. This is probably because they are used for engaging in a dialogue and stance-taking with the content, rather than with other users, as revealed in the analysis below. Before the elaboration of the communicative function of both keywords, the patterns of use of *think* and *agree* in the FL corpus that this analysis based on are first described.

In the analysis, only *I think* and *I agree* are examined because they are the most frequent bigrams of the two independent keywords. According to log-likelihood ratio test, *I think* and *I agree* are also used significantly more often in the independent posts than the initiating posts. Of the

¹²⁵<https://www.futurelearn.com/courses/learn-to-code/1/comments/8953616>

16682 occurrences of *think* in the independent posts, 68% are in the bigram *I think*. Another 2% are in the pattern *I + adverbs + think*, including *I also think*, *I still think*, *I really think*, *I definitely think*, *I personally think*, *I always think*. The adverbs could be a booster to the meaning of *I think*. There are also 7% of occurrences of *think* in negation acts, for example, *I don't/do not think*, *I didn't/did not think*, *I wouldn't think*, *I can't/cannot think*, *I couldn't think*. Because neither the adverb pattern nor negation occur frequently, only *I think* is examined.

Of the 2318 occurrences of *agree* in the independent posts, 59% are in the bigram *I agree*. Another 15% are in the pattern *I + adverbs + agree*, including *I totally agree*, *I do agree*, *I also agree*, *I completely agree*, *I would agree*, *I absolutely agree*. As with *I think*, the adverbs could be a booster to the meaning of *I agree*. There are also negations for *I agree*, as in *I don't/do not agree*, but they account for only 2% of the occurrences of *agree*. Because the negation only appears rarely, it can be safely assumed that almost all the time when *agree* is used in the independent posts, it is used to express agreement rather than disagreement. Another independent keyword *yes* also signals agreement.

7.3.1 *I think*

I think is typically used to express user's certainty or commitment to the truth value of a proposition, thus marking their subjectivity, rather than bare assertions (Kärkkäinen, 2003). However, the nuanced discourse function of *I think* depends on the textual and situational context, such as the existence of boosters or hedges in the utterances, tone of voice, and power dynamics between interlocutors (Baumgarten & House, 2010; Kärkkäinen, 2003; Pöldvere et al., 2016; Simon-Vandenberg, 2000). In an online political discussion, users seem to use *I think* to talk about issues directly affecting them (Sotillo & Wang-Gempp, 2016). In a similar vein, in the current context of FutureLearn discussions, it is found that *I think* is used for talking about their personal learning, as well as taking stances in response to prompts or course content.

7.3.1.1 Personal learning

Users comment on their learning journey with *I think* together with collocates such as *will* (n=416), *is* (n=2462), *are* (n=754) which are on the cline of certainty and assertion. In these situations, users express how they feel about the course or course materials (e.g., “...*I think* that this course will help me a lot.”¹²⁶; “*I think* I get full contract management articles from this course...”¹²⁷) or evaluate the proposition raised in the course content (e.g., “*I think* this is a really good option to remediation of soils...”¹²⁸) or the learning activities (e.g., “*I think* this assessment is fun!...”¹²⁹), suggesting that users not only passively go through course content but actively engage in reflective learning.

7.3.1.2 Stance-taking

In the independent posts that express stance, *I think* is used for taking strong stances or making speculations. The rhetorical force of *I think* depends on the textual context, consistent with previous findings (Kärkkäinen, 2003; Pöldvere et al., 2016; Simon-Vandenberg, 2000). The propositions introduced by *I think* tend to be stronger in stance when it is a generalization with the use of *we* (n=638), as shown in “*I think we should* be open to the possibility of life elsewhere...”¹³⁰, or when *I think* collocates with words such as *very* (n=455), *important* (n=361), *should* (n=582), as shown in “*I think it is very important* to build up a trusting relationship...”¹³¹. In contrast, *I think* is used for hedging purposes and stating speculations, when it collocates with modals such as *can* (n=455), *would* (n=455), as shown in “...However, *I think* that technology *can* become more troublesome when the pupils do not understand the expectations with it...”¹³².

Users also tend to use *I think* to take a stance when they respond to discussion prompts that ask for their experiences or opinions (Figure 7.1), especially when the phrase *do you think* is used in

¹²⁶ <https://www.futurelearn.com/courses/dyslexia/1/comments/4511175>

¹²⁷ <https://www.futurelearn.com/courses/contract-management/4/comments/18425308>

¹²⁸ <https://www.futurelearn.com/courses/soils/1/comments/6373234>

¹²⁹ <https://www.futurelearn.com/courses/corpus-linguistics/1/comments/354199>

¹³⁰ <https://www.futurelearn.com/courses/moons/1/comments/529114>

¹³¹ <https://www.futurelearn.com/courses/palliative/1/comments/17202046>

¹³² <https://www.futurelearn.com/courses/dyslexia/1/comments/4704775>

the prompt (Figure 7.2). Given that it is a response to a question, it might not be framed in a way that opens up the dialogic space to others, while other users may also tend to reply to the question, rather than replying to others' answers. These prompt-focused responses are also expressed with *I agree*, which is discussed next.

Figure 7.1 Independent post 4626201

Source: <https://www.futurelearn.com/courses/dyslexia/1/comments/4626201>

Discussion Prompt: Should students with specific learning differences learn foreign or additional languages? Explain your answer.

Independent Post 2015-04-27 18:31:20 Like: 1

User d1-4143

I think that all children should have the opportunity to learn another language. The possible cognitive benefits associated with being bilingual may support children in overcoming some of the challenges associated with their learning difference.

Note. "*I think*" is used to respond to discussion prompt.

Figure 7.2 Independent post 17474926.

Source: <https://www.futurelearn.com/courses/palliative/1/comments/17474926>

Discussion Prompt: Think about how hard it would be for these patients if they did not have family support. *Do you think* they would have the same outcome?

Independent Post 2016-11-05 22:28:59 Like: 0

User p1-1279

I think that palliative home care aren't feasible without a strong family . It's because, also if the operators go patients home twice or more a day, this organization can't provide all the needs of the patient. It's very difficult understanding if and when the patient asingle must go to the hospice or other resident home.

Note "*I think*" may be used to respond to the "*Do you think*" discussion prompt. The label "*Discussion Prompt*" is mine and does not show on the course step.

7.3.2 *I agree*

I agree is used to align with what has been mentioned in the communicative context (Du Bois, 2007).

In the independent posts, *I agree* is used typically at the start of a post while users further elaborate on their agreement (e.g., “*I agree with [Name]. The will is there but ...*” in Figure 7.4 and other examples in Figure 7.3, Figure 7.5, 7.6). There are also a few occasions where *I agree* is used for mere expression of agreement without any expansion (e.g., “*I agree with Ovid.*” in Figure 7.7 and other examples in Figure 7.8, 7.9, 7.10). *I agree* have been used to align with course content (e.g., “*I agree with identifying a champion...*” in Figure 7.3), other users (e.g., “*I agree with [Name]...*” in Figure 7.4), or the general trend in the discussion spaces (e.g., “*...I agree with all things listed by others...*” in Figure 7.5). Similar to *I think*, sometimes *I agree* is also used in response to a discussion prompt that asks for their stances, especially in the phrase “Do you agree...” (Figure 7.10). The expression of agreement in the independent posts may be indicative of users’ engagement and responsiveness with the materials and others’ comments (Bolander, 2012), as explained below.

7.3.2.1 *I agree with elaboration*

In the expression of agreement with elaboration, users reformulate (e.g., “*I agree ... so it enables...*” in Figure 7.3), summarize (e.g., “*I agree with [Name]. The will is there but ...*” in Figure 7.4) or expand on what others have said (e.g., “*I agree with all things listed by others. I add a few...*” in Figure 7.5). This is consistent with Baym's (1996) conclusion that the expression of agreement and additional elements are included by Usenet users to make one’s own novel contributions relevant to the ongoing discussion. This interpretation is especially relevant when the agreement is expressed towards a specific user by creating a new post, instead of by replying to the target user, for example “*I agree with [Name]...*” in Figure 7.4, if it is assumed that posts are new contributions and replies are subordinated to others’ posts in online discussions (Ksiazek & Lessard, 2016). Similarly, if a user agrees with comments contributed by more than one user, for example “*I agree with all things listed by others*” in Figure 7.5, such that replying to every single user is not possible, creating a new post might be more feasible. More importantly, *I agree* in these independent posts may serve to signal a

post's relevance to the ongoing discussion or to acknowledge others for the purposes of discourse organization and community building, before the users embark on their own opinions (Baym, 1996; Bolander, 2012; Lapadat, 2007; O'Keeffe & Walsh, 2012). Thus, the practice of posting agreement with elaboration, instead of replying to a post, can be interpreted as broadcasting one's new contributions to all other users while engaging in the current communicative context.

Figure 7.3 .Expression of agreement towards course content with elaboration.

Source: <https://www.futurelearn.com/courses/digital-accessibility/2/comments/19737567>

Independent post 2017-02-15 Like: 2

User a2-571

I agree with identifying a champion on disability issues and organizations promoting disability within the workplace with courses such as disability awareness so it enables people to gain more of an insight into how much of an impact disabilities can have.....

[In the video in the step, the speaker gives an example of an organisation promoting disability champions.]

Figure 7.4 Expression of agreement towards a specific user with elaboration.

Source: <https://www.futurelearn.com/courses/palliative/1/comments/16740444>

Independent post 2016-10-18 11:09:01 Like: 1

User p1-787

I agree with [Name]. The will is there but palliative care services and hospices are constrained by lack of resources and funding / coordination of a wide variety of organisations, all of which have their own constraints. (I am in the UK as well)

[This user agrees with a post created by the named user right before them.]

Figure 7.5 Expression of agreement with elaboration towards comments contributed by other users.

Source: <https://www.futurelearn.com/courses/digital-accessibility/2/comments/19766781>

Independent post 2017-02-15 Like: 1

User a2-422

I downloaded the .odt file. *I agree* with all things listed by others. I add a few: - No title - Often in an article, we need to provide something like "Figure 1 description" under an image/figure; "Table 1 description" above a table. - Abbreviations not described at the first time it occurs. - Two font styles in texts - Table does not have heading, difficult for reading, print.

[There is a discussion prompt in this step "How many violations of the accessibility guidelines can you spot?"]

Besides posting agreement towards others, users also post their agreement towards course content, especially discussion prompts. This can be construed as user-content interactions. The discussion prompts can also be perceived as course educators addressing all users. It may thus give the impression that users have to create their own posts in response to the discussion prompt, for example “Do you agree” in Figure 7.6, as well as “Do you think” in Figure 7.2. Therefore, the fact that *I agree*, as well as *I think* are used more often in the independent posts than in the initiating posts may be due to the possibility that the users are engaging with course content rather than with other users. This observation is in line with the aim of the discussion prompts in FutureLearn that encourages users to engage in conversations with oneself or the course content (R. Ferguson & Sharples, 2014).

Figure 7.6 “I agree” may be used to respond to the “Do you agree” discussion prompt.

Source: <https://www.futurelearn.com/courses/inequalities-in-personal-finance/1/comments/4325879>

Discussion Prompt: The video you have just watched reports that many governments are saying:

- state pensions are unaffordable and must be cut back
- state pensions place an unfair burden on younger generations
- everyone must make their own private savings for retirement as well
- people are not saving enough – they must put more into their private pensions.

Do you agree with these statements? Note your views here and then look back at the end of this week to see if they have changed.

Independent post 2015-03-26 16:40:09 Like: 6

User f1-95

First of all whilst *I agree* that the younger population are burdened by the state pension, I don't think it would be wise to cut back on it because retired persons need a certain amount of money for their basic needs.

Lastly, the fact that independent posts expressing agreement are less likely to receive replies can be due to the organization of agreement discourse (Pomerantz, 1984). According to Pomerantz's investigation of oral conversation, an assessment of an issue is first raised, followed by a second assessment, either agreement or disagreement. Generally, one agreement is sufficient for the second assessment, as a third assessment is not preferred unless there is novel content to be added or a disagreement. In the current setting, agreement in the independent posts is the second assessment, i.e., alignment, to the course content, discussion prompts or other users' posts, which is the first assessment of an issue. Typically, no preferred third assessment, i.e., reply, is required further in the sequence. Therefore, an agreement post is less likely to receive replies, thus there are more agreement expressions in the independent posts than in the initiating posts, as revealed by the keyword analysis.

7.3.2.2 Mere expression of agreement / agree

Beside agreement with elaboration, there are also independent posts that only express agreement without further expansion, as shown in Figure 7.7, 7.8, 7.9, 7.10. Similar to the agreement with elaboration, it is used to align with course content (Figure 7.7 and Figure 7.8), other users' posts (Figure 7.9) and in response to discussion prompts (Figure 7.10).

Figure 7.7 Mere expression of agreement towards course content.

Source: <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/19142339>

Independent post 2017-01-26 04:44:04 Like: 0

User n4-3358

I agree with all this information.

[The step content is an article describing Eatwell Guide with subheadings "How has it changed?", "Who does it apply to?", "How to use it", "What it tells us".]

Figure 7.8 Mere expression of agreement towards course content.

Source: <https://www.futurelearn.com/courses/ancient-health/1/comments/19724777>

Independent post 2017-02-14 16:52:04 Like: 0

User ah1-758

I agree with Ovid.

[Ovid is mentioned in the step content “In his *Art of Love*, Ovid includes dealing with nostril hair and clipping nails as acceptable for men (1.505–524)”.]

Figure 7.9 Mere expression of agreement towards comments contributed by others.

Source: <https://www.futurelearn.com/courses/digital-accessibility/2/comments/20405894>

Independent post 2017-03-08 Like: 0

User a2-275

I agree tolerance.

[*Tolerance* have been mentioned by three other users before.]

Figure 7.10 Mere expression of agreement towards comments contributed by others.

Source: <https://www.futurelearn.com/courses/dyslexia/1/comments/4690476>

Independent post 2015-05-02 10:24:24 Like: 1

User d1-3626

I agree with the comments posted. Faced the exact same difficulties.

[There is a discussion prompt in this step “Note down how the activity made you feel and why you felt like that. Make a list of the strategies you used to accomplish it. Consider what this experience tells you about how students with learning difficulties might feel when facing the task of learning another language.”]

These mere expressions of agreement would not receive replies because there is no concrete substance in it for others to reply to, and possibly it is the second assessment that does not

require further alignment (Pomerantz, 1984). However, similar to agreement posts which contain elaboration, these posts could be indicative of users' interaction and engagement with the course, perhaps to some extent an overt expression of vicarious learning (Mayes, 2015). Users do not expand on their agreement perhaps because they would like to avoid redundancy or repeating what has been stated by others (Baym, 1996).

7.4 Discussion

Based on the functional grouping and analysis of *I think* and *I agree*, two general patterns of discourse practices engaged in by users in the independent posts can be identified: 1) expressions of appreciation and positive sentiments regarding the course and their own learning; 2) stance-taking in response to course content or discussion prompts. These discourse practices are oriented towards oneself, course content or educators, in contrast to information sharing/seeking and stance-taking in the initiating posts that orient towards potential conversational partners. This finding thus addresses the RQ1 regarding the differences between independent posts and initiating posts.

It can be argued that the discourse practices underlying these general patterns may render the independent posts less likely to invite replies from other users in the online discussions, compared to those found in the initiating posts, as identified in Chapter 6. One may argue that probably these independent posts are not seen by others. However, as can be seen by some of the examples, they receive likes from others, suggesting that they are seen but the discourse do not establish a dialogic space for others to reply (See Appendix E for numbers of likes received by independent posts).

In this section, two discourse practices are identified to explain why the independent posts are less likely to receive replies from the perspective of dialogic space and intersubjectivity. At the same time, I argue that, although the discourse practices found in the independent posts are less likely to start user-user interactions, they indicate user-content interactions in the online discussions

(Ksiazek & Lessard, 2016; Ziegele et al., 2014), or conversation with oneself or the content, as encouraged by the design of Futurelearn and the discussion prompts in some specific cases (R. Ferguson & Sharples, 2014; Laurillard, 2012).

Sharing one's appreciation and epistemic status of "becoming knowing"

The words of appreciation and positive sentiments in the independent posts are expressed through keywords categorised as positive evaluations, emotional and attitude verbs, quantifiers and boosters, meta-language, first person pronouns, politeness formulae and the exclamation mark (!). Comments on positive learning experience are also prominent in the independent posts, as evidenced by the expressions of appreciation, meta-language and epistemic verbs found to be independent keywords. It is likely that these discourse practices are less likely to receive replies because these positive expressions are addressed towards course facilitators and designers, such that other users are not the target interlocutors to reply. These could be considered feedback on the course from the users. Furthermore, words of appreciation and reflection on positive learning experience may be for sharing purposes rather than inviting responses, similar to the hotel reviewers who wrote about their hotel stay for sharing purposes rather than engaging others for alternative viewpoints (Tian, 2013). This contrast with discourse practices found in the initiating posts that set up a dialogic space for other possible voices. The sharing of appreciation and positive learning experience also differs from the sharing of information that attracts other users to reply with thank-you messages, as shown in the information sharing practices found in the initiating posts in Chapter 6. Lastly, the sharing of positive learning experience may be an indication of "becoming knowing", that does not require others to fill in any knowledge gap, compared to the discourse practices in the initiating posts of expressing "unknowing" and "partially knowing" status that invite other users to fill in the gap.

Establishing dialogic relationship with prior utterances rather than inviting responses

Although the appreciation and reflection on positive learning experience may not be designed to invite replies, they can still be considered dialogic because they are expressed towards

the course facilitators and designers or in response to the discussion prompts that ask them to reflect on their learning. Similarly, when users take stances with phrases such as *I think* and *I agree* in response to discussion prompts, such as “Do you think...” or “Do you agree...”, they are engaging in a dialogue with educators, assuming that the prompts are perceived as raised by educators. Also, where users’ posts are designed as answers towards the questions in the discussion prompts, it is likely that they are not simultaneously designed to invite others’ views. These posts refer to the prior utterances rather than inviting future utterances, compared to the initiating posts which address potential users. Therefore, it can be argued that the dialogic nature of the discourse practices in the independent posts is retrospective, whereas in the initiating posts it is prospective.

Furthermore, responding to a discussion prompt resembles a question and answer sequence such that no further reply is needed. It could be the case that course content, discussion prompts or other users’ comments are deemed as the first assessment of an issue, while users’ agreements in the independent posts are the second assessment that typically does not require further alignment, thus no replies (Baym, 1996; Pomerantz, 1984). This contrasts with the negation acts in the initiating posts that normally set up an opposing voice among other users for further interactions.

These two discourse practices are less likely to receive replies because they are directed towards prior utterances in the communicative context, for example, discussion prompts, course content, other users’ comments, or because they are words of appreciation that are not designed to engage in alternative viewpoints. However, both these discourse practices establish a dialogic relationship either towards the educators in the case of appreciation, or course content in the case of taking a stance, suggesting that users have engaged with the content of the course. Users’ engagement with the course content is further evidenced by the observation that users’ use of *agree* and *think* in response to the discussion prompts containing the same words.

7.5 Conclusion

This chapter turned our attention from initiating posts to independent posts, extending the investigation of the potential start of a thread to the perspective of what makes a post less likely to receive replies. This complements the existing literature and Chapter 6 that focus only on what makes a post more likely to receive replies. The findings in this chapter further contribute to our understanding of the dialogic nature of online discourse in the sense that the dialogue can be towards the page content, rather than towards other users. Because the dialogic relationship is established retrospectively rather than prospectively, discourse practices in the independent posts can therefore be considered as contracting dialogic space. This thus addresses RQ2 regarding how users' discourse practices may not initiate dialogic conversations.

Importantly, the analysis of independent posts also contributes to our understanding of user-content interactions in online spaces. The discourse practices in independent posts, as well as their overwhelming numbers, can indicate successful engagement by the hosts of the websites (Preece & Maloney-Krichmar, 2005; Sharples & Ferguson, 2019). Specifically, the engagement as overtly realized by the discourse practices attests to the design of FutureLearn, i.e., to encourage conversation with oneself while interacting with the course content. User-content interactions are also common in commenting spaces in news websites (Ksiazek & Lessard, 2016), blogs (Bolander, 2012), and YouTube (Herring, 2013), probably because all these platforms have the similar design of "discussion in context", rather than stand-alone discussion forums. As has been concluded in these previous studies, this design could perpetuate prompt-focused behaviour. Therefore, the finding in this chapter also reflects the fact that users' behaviour is influenced by technology and the design of the online space too.

Additionally, this finding also suggests that some users might not necessarily write a post with the aim to invite responses from others, or more generally participate in the online discussion for socialization or deliberation with others, as found in previous studies (Delahunty, 2018; Pendry & Salvatore, 2015). As will be shown later in Chapter 8, some users even decide to "agree to disagree"

to disengage from a conversation after others reply and disalign with their stance. This indicates that some users may be participating in online discussions, not for social interactions, but simply to engage with content while also contributing by posting or liking. This is also supported by the observations in Chapter 5 that most users contribute more posts than replies, or simply just post, without returning to the same thread to continue conversations. This is likely the case for some of the FutureLearn users given that it is a MOOC platform aimed mainly at learning, despite its social learning design. As argued in Chapter 3, online discussions create an equal ground, such that users are not obliged to engage in user-user interactions but can choose to interact with the platform and express their views only (Cavanagh, 2007; Herring, 2013).

However, as also argued earlier, a dialogic conversation not only requires multiple voices to be raised and heard, but also for these voices to be entertained and deliberated through processes of intersubjectivity. Although independent posts have been read or even liked by others, without user-user interactions, intersubjectivity is not possibly to take place, not to mention a dialogic conversation. This can potentially compromise the online deliberation in online discussions (Freelon, 2015; Friess & Eilders, 2015; Preece & Maloney-Krichmar, 2005). In contrast, initiating posts, which start user-user interactions, can potentially allow a dialogic conversation to arise, as we shall see in the next chapter where users' replies are analysed.

Admittedly, there are also independent posts which are designed to start user-user interactions but to no avail (Burke & Kraut, 2008), probably due to situational factors including the order in which the post appears in relation to other posts (Hewitt, 2003; Jeong & Frazier, 2008) and size of the discussions (Himmelboim, 2008), as well as the prompt-focused behaviour that reduces the number of users that reply (Bou-Franch & Garcés-Conejos Blitvich, 2014; Herring, 2013). As argued in Chapter 1, these factors are beyond users' control while this thesis focuses on users' agency in establishing social relationship with their discourse.

In short, this analysis of discourse practices in the independent posts and initiating posts in this chapter and Chapter 6 may have different implications for different users, depending on their

goals of participating in an online space, i.e., socializing and deliberating with others or merely engaging with the course content. Users keen on engaging in user-user interactions might want to avoid discourse practices found in the independent posts in this chapter because the practices may render their posts less likely to receive replies from other users in the online discussions. Instead, they might want to utilize the discourse practices found in the initiating posts as identified in Chapter 6 to increase their chance of receiving replies from others. Otherwise, users' discourse in the independent posts may reveal their user-content interactions in FutureLearn, thereby indirectly informing course designers and educators of users' learning experience.

Chapter 8

A keyword analysis of replies: Discourse practices for intersubjectivity

8.1 Introduction

Moving on from the potential starts of conversations, this chapter explores how dialogic conversations can develop in online discussions by examining replies posted underneath initiating posts. Replies are assumed to indicate explicit user-user interactions because they are addressed towards the initiating post or to replies that are posted before them within the same thread (Ksiazek & Lessard, 2016; Lewis, 2005). Therefore, analysis of replies illuminates how user-user interactions unfold, and how they might be sustained and developed into a dialogic conversation.

As reviewed in Chapter 3, previous studies have revealed various social and discourse practices within threads, including metapragmatic discussions (Kleinke & Bos, 2015; Tanskanen, 2007), second story (Kääntä & Lehtinen, 2016), identity performance (Grabill & Pigg, 2012; Jaworska, 2018; Stommel & Koole, 2010), negotiation (Littleton & Whitelock, 2005; Ziegler et al., 2014), and agreement and disagreement (Baym, 1996; Bou-Franch & Garcés-Conejos Blitvich, 2014). Informed by these findings, the focus of this investigation is not decided a priori but derived initially from word frequency and keyword analysis. Micro-analysis of selected sustained threads is then conducted to illustrate how the discourse practices derived from the keyword analysis unfold in the threads and facilitate intersubjectivity. In other words, the investigation proceeds from bottom up – from the discourse practices evident in the threads to what users are trying to achieve through their language use. The chapter thus addresses the first two research questions:

RQ1: What are the differences in linguistic features and discourse practices that regularly occur in replies compared to the initiating posts and independent posts?

RQ2: How do these discourse practices initiate, sustain or hinder dialogic conversations in online discussions?

To achieve this, the chapter starts with the keyword analysis that reveals words that are used significantly more often in replies than in initiating posts and independent posts. The distribution of these keywords in different types of replies and threads are also examined to further understand their function in the dynamics of user-user interactions, especially to compare the replies between short and long threads, given that rarely occurring long threads are more likely for negotiation and intersubjectivity to take place, as argued in Chapter 3. The reply keywords are then grouped based on their communicative functions in context.

Micro-analysis of three threads in which users engage with opposing stances is then conducted to explore how user-user interactions unfold via the discourse practices derived from the reply keywords, thus revealing stance-taking and processes of intersubjectivity in online discussions. The micro-analysis also highlights the distinctive features of dialogic conversations. An analysis of *agree to disagree/differ* is then conducted to follow up on how users manage unresolvable disagreement, given the importance of disagreement in online discussions, as argued in Chapter 3. This chapter concludes by arguing that disagreement, although undesirable, can be a constructive dialogic space where intersubjectivity and multiple voices are explored with discourse practices such as concession and meta-language.

8.2 Keyword analysis

The keyword analysis for replies is a two-step process. In the first step, two comparisons are made: replies vs. initiating posts and replies vs. independent posts (see section 4.5.3 for the statistical procedure). In the second step, the keywords of replies that are found from both comparisons are extracted as the reply keywords. The resulting reply keywords are thus those used significantly more often in replies than in both initiating posts and independent posts.

There are 63 keywords found to be used more often in replies compared to initiating posts (see Appendix F for the full statistics of log-likelihood ratio test and dispersion measure) and 178 keywords found to be used more often in replies compared to independent posts (see Appendix G for the full statistics of log-likelihood ratio test and dispersion measure). Among these keywords, 57 keywords are common across these two comparisons. They are thus the reply keywords that are used significantly more often when compared to both initiating posts and independent posts (Table 8.1), and will be investigated for their communicative functions in replies. It is worth mentioning that the results of the two comparisons by themselves could be of significance but are not explored further because this thesis only focuses on the discourse of replies that are significantly different from both types of posts.

Table 8.1 Reply keywords that are used significantly more often when compared to both initiating posts and independent posts.

Keyword	Normalized Frequency ² in Replies	Normalized Frequency in Initiating Posts	Normalized Frequency in Independent Posts	Effect size when vs. initiating posts	Effect size when vs. independent posts	Average effect size ³
reply	16.5	1.9	0.8	8.82	21.22	15.02
jane ¹	6.6	0.5	0.5	14.45	14.57	14.51
michael	9.2	1.2	1	7.34	8.84	8.09
ah	7.3	1.3	1	5.64	7.60	6.62
agree	160	26.6	37.6	6.01	4.25	5.13
posting	5	1.3	0.8	3.90	6.08	4.99
yes	117.3	20.2	30.6	5.79	3.84	4.82
:-	27.3	6.6	6.6	4.14	4.12	4.13
agreed	9.3	2.4	3.2	3.93	2.95	3.44
john	12.7	5	3	2.55	4.20	3.37
your	245.6	75	73.2	3.27	3.36	3.32
thanks	174.2	45.2	65.1	3.85	2.67	3.26
comment	33.9	13.6	9.3	2.48	3.63	3.06
link	54	22.5	15.3	2.40	3.53	2.97
exactly	21.1	9.4	6.9	2.25	3.08	2.66
oh	15.1	7.2	4.7	2.11	3.21	2.66
luck	12.1	4.2	5	2.87	2.40	2.64
you	855.2	340.3	322.2	2.51	2.65	2.58
sorry	28.4	15.7	8.6	1.80	3.31	2.56
're	41.6	19.6	14.9	2.13	2.79	2.46
totally	22.8	8.8	9.8	2.58	2.31	2.45
hi	102.5	28.8	84.4	3.55	1.21	2.38
true	33.7	16.9	12.9	1.99	2.62	2.31
post	21.7	11.5	8.4	1.88	2.57	2.22
absolutely	14.6	6.4	7.6	2.29	1.92	2.10
indeed	21.7	11	9.9	1.98	2.20	2.09
point	74.7	41	33.5	1.82	2.23	2.03
mine	13.1	6.4	7	2.05	1.89	1.97
thank	117.7	49.3	83.7	2.39	1.41	1.90
above	30	19.2	13.6	1.57	2.21	1.89
too	167.7	89.6	95.4	1.87	1.76	1.81
'll	37.8	20.6	21.4	1.84	1.77	1.80
&	76	39.9	46.1	1.90	1.65	1.78
maybe	58.5	36.3	31	1.61	1.89	1.75
same	129	88.4	72.2	1.46	1.79	1.62
right	80.4	53.1	52.1	1.51	1.54	1.53
...	249.5	182.2	149.1	1.37	1.67	1.52
said	49.7	36.2	30.4	1.37	1.64	1.51
!	634.8	374.9	496.4	1.69	1.28	1.49
say	80.4	61.5	48.4	1.31	1.66	1.48
probably	48	35	31.1	1.37	1.54	1.46

's	461.3	344.7	302.7	1.34	1.52	1.43
just	214.4	163.2	140.7	1.31	1.52	1.42
n't	433.2	345.8	296.5	1.25	1.46	1.36
-	354.7	306.2	231.5	1.16	1.53	1.35
did	147.8	124.1	100.6	1.19	1.47	1.33
go	76.3	59.5	55.9	1.28	1.37	1.32
;	222.2	192.5	149.7	1.15	1.48	1.32
if	361.2	308.5	248.4	1.17	1.45	1.31
no	157.5	131.5	118	1.20	1.34	1.27
those	97.3	77.8	77.2	1.25	1.26	1.25
had	207.7	178.2	164.5	1.17	1.26	1.21
do	422.4	364.6	336.6	1.16	1.25	1.21
that	1258.1	1105	1033.2	1.14	1.22	1.18
it	1330.7	1142.4	1128.2	1.16	1.18	1.17
people	215.6	186	186.5	1.16	1.16	1.16
but	568.8	517.1	496.7	1.10	1.15	1.12

Note. Only normalized frequency and effect size are presented here to reveal the differences between the replies and the two types of posts. The full statistics for the log-likelihood ratio test and dispersion measure (i.e., the criteria to decide whether a word is a keyword) of each reply keyword can be found in Appendix F and G.

¹ Names are not anonymized here because specific individual users are not identifiable this way. The same name can be used to refer to different users. For example, both Jane and Michael appear in all MOOCs except accessibility-2, while John appears in all MOOCs, and so almost certainly refer to multiple users. Although specific names are found to be statistically significant, it might not have empirical significance. Rather, a more possible conclusion is that users address each other with names in their replies, and some names are more common than others.

² Normalized frequency is in per 100,000 words *and is rounded so there is no decimal places*

³ The keywords are ordered by average effect size of the two comparisons. The effect size is measured by relative risk, that is the ratio of the normalized frequency in replies to the normalized frequency in the two types of posts. The calculation is based on unrounded normalized frequency rather than the rounded one shown here.

The effect size and word frequency of the reply keywords vary, as shown in Table 8.1. The effect size of the top seven reply keywords have a relatively large effect size (ranging from 4 to 14), suggesting the nature of replies might be very different from either type of posts. From the top few reply keywords, it can be argued that replies are more interactive than both types of posts, as evidenced by names which are used to address a specific user, *yes*, *agree*, and *agreed* which are used to align with others' position in stance-taking, *reply*, *posting* and *your* to refer to others' comments, *ah*, *:-*, *hi*, *thank*, *thanks* to respond to others.

A few reply keywords are also found to have significantly different distribution across the different types of replies (Table 8.2, Appendix H for full statistics) and threads of different lengths (Table 8.3 and Table 8.4, Appendix I and Appendix J for full statistics). As shown in Table 8.2, *agree*,

too, *same* are used significantly more often in users' first contributions as replies in a thread compared to subsequent contributions, suggesting users typically express agreement or similarity in ideas, experience, problems and situations in their first contributions in a thread, but are less likely to leave such a response once they have already contributed to the thread. Instead, when they come back to a thread that they have contributed to before, their replies are typically responsive and interactive, as evidenced by the keywords *reply*, *thank*, *ah*, *thanks*, *oh*, *yes*, *you*, *sorry*, and *but* (Biber et al., 1999). *Sorry* and *but* may be used in disagreement (Baker, 2014; Baym, 1996), and will be explored later in the micro-analysis.

Table 8.2 Reply keywords used significantly more often in first contributions and subsequent contributions in a thread.

	Normalized frequency ¹ of first contributions	Normalized frequency of subsequent contributions	Effect size ²
<u>Keywords used more frequently in the first contributions</u>			
agree	200	81	2.47
&	88	52	1.71
same	146	96	1.52
too	187	130	1.44
<u>Keywords used more frequently in the subsequent contributions</u>			
reply	6	38	6.90
thank	60	231	3.83
ah	4	14	3.57
thanks	94	333	3.56
sorry	16	52	3.23
oh	11	24	2.29
'll	28	57	2.01
:-	22	38	1.73
yes	106	140	1.32
...	232	284	1.23
but	540	627	1.16
you	825	915	1.11

Note. Only normalized frequency and effect size are presented to reveal the differences between the two types of replies. The full statistics for the log-likelihood ratio test and dispersion measure (i.e., the criteria to decide whether a word is a keyword) of each keyword can be found in Appendix H.

¹Normalized frequency is in per 100,000 words and is rounded so there is no decimal places.

²The effect size is measured by relative risk, that is the ratio of the normalized frequency in first contributions to the normalized frequency in subsequent contributions. The calculation is based on unrounded normalized frequency rather than the rounded one shown here.

Table 8.3 Reply keywords used significantly more often in short threads.

	Normalized frequency ¹ in short threads	Normalized frequency in sustained threads	Effect size ²
agree	200	102	1.96
yes	136	91	1.50
hi	116	83	1.40
too	189	137	1.38
same	144	108	1.34
!	702	538	1.30
you	898	793	1.13

Note. Only normalized frequency and effect size are presented to reveal the differences between the replies in short threads and sustained threads. The full statistics for the log-likelihood ratio test and dispersion measure (i.e., the criteria to decide whether a word is a keyword) of each keyword can be found in Appendix I.

¹The normalized frequency is measured by per 100,000 words and is rounded so there is no decimal places.

²The effect size is measured by relative risk, that is the ratio of the normalized frequency in the replies in short threads to the normalized frequency in the replies in sustained threads. The calculation is based on unrounded normalized frequency rather than the rounded one shown here.

Table 8.4 Reply keywords used significantly more often in the first reply of one-reply threads than those of threads with more than one reply.

	Normalized frequency ¹ in the first reply of one-reply threads	Normalized frequency in the first reply of threads with more than one reply	Effect size ²
&	109.91	63.06	1.74
agree	333.67	202.16	1.65
yes	162.51	101.58	1.60
!	761.54	585.52	1.30

Note. Only normalized frequency and effect size are shown to reveal the differences between the two types of first reply. The full statistics for the log-likelihood ratio test and dispersion measure (i.e., the criteria to decide whether a word is a keyword) of each keyword can be found in Appendix J.

¹The normalized frequency is measured by per 100,000 words and is rounded so there is no decimal places.

²The effect size is measured by relative risk, that is the ratio of the normalized frequency in the first reply of one-reply threads to the normalized frequency in the first reply of threads with more than one reply. The calculation is based on unrounded normalized frequency rather than the rounded one shown here.

Surprisingly, none of the reply keywords are used significantly more often in threads with five or more than five replies (i.e., long threads), suggesting either that there might not be any word or expression that can increase the chance of sustaining a thread, or that discourse practices sustaining a thread are not realized by particular keywords or expressions. Micro-analysis of sustained threads in section 8.5 will complement this null result of the keyword analysis to improve our understanding of discourse practices that might be facilitative of the development of a thread. In contrast, as shown in Table 8.3, it is found that *agree*, *yes*, *too*, *same*, *hi* are used more often in the threads with fewer than five replies (i.e., short threads), suggesting that expressions of agreement might not lead to sustained conversations, although it creates a supportive space and indicates users' engagement with each other. This could be because other users may not have further content to add in an agreeable situation, and typically there are no more than two expressions of agreement within a conversation (Baym, 1996; Pomerantz, 1984). Lastly, as revealed in Chapter 5, more than half of the threads consist of one reply, so a comparison is thus made between the first reply in one-reply threads and those in threads with more than one reply to examine if the first reply to an

initiating post may also have an effect on whether the threads will continue (Table 8.4). Similar to the previous two comparisons, keywords used for expressing agreement, *agree* and *yes* are found used more often in one-reply threads, confirming again that agreement may not necessarily lead to sustained interactions or more contributions. Exclamation mark (!) is also used more often in the reply of one-reply threads, probably because users use it to express their agreement or emotion towards the initiating post, and this reply does not further develop the interactions.

8.3 Functional grouping of reply keywords

The functional grouping of the reply keywords based on their communicative functions is presented in Table 8.5. Most of these keywords indicate three general patterns of discourse practice: alignment/disalignment in stance-taking; meta-language on others' comments; and interactive language, all of which attest to the user-user interactions in the replies. In the elaboration below, no example replies with the keywords are presented because the discourse practices in replies are best illustrated via micro-analysis of the threads, which will be presented in section 8.5. This section only aims to give an overview of the discourse in the replies based on the keywords.

Table 8.5 Functional grouping of the reply keywords.

Functional grouping	Reply keywords
Hedges (8.3.1) ¹	maybe, probably
Boosters (8.3.1)	exactly, absolutely, totally, indeed, too, just
Evaluatives (8.3.1)	true, right, same
Negation (8.3.1)	n't, no
Attitude verbs (8.3.1)	agree, agreed
Communicative verbs (8.3.2)	say, said
Meta-language on discussions (8.3.2)	reply, posting, comment, link, post, point
Polite speech-act formulae (8.3.1 & 8.3.3)	yes, thanks, thank, hi, sorry, luck ²
Interjections (8.3.3)	ah, oh
Pronouns (8.3.1 & 8.3.2)	you, your, mine, it, those
Connectors (8.3.1)	but, if
Grammatical particles	that, 's, 're, do, had, did, above, 'll ³
Punctuations	!, &, ..., ;, -
Names (8.3.3)	jane, michael, john
Uncategorized	people, go, :- ⁴

¹ The number in the bracket indicates the subsections in the text that the keywords in the group are described. The subsections also correspond to the three general patterns.

² 63% of *luck* follow *good* to form *good luck*.

³ 'll is the only modal, so it is categorized as grammatical for convenience.

⁴ 99% of :- is followed by) to form emoticon :-).

8.3.1 Stance-taking

Much stance-taking in replies takes the form of alignment/agreement or disalignment/disagreement with what has been said in the initiating posts or other replies, as evidenced by the following functional groups of keywords. Boosters (*exactly, absolutely, totally, indeed*), evaluatives (*true, right*), attitude verbs (*agree, agreed*), especially in the form of *agree with you/your...*, and polite speech-act formulae (*yes*), are typically used for alignment. In contrast, negations (*no, n't*) are used mostly for disalignment, especially in the collocation of *n't... agree/think*. *Sorry* and the collocation *sorry...but* are also found to preface disagreement (Baker, 2014; Baym, 1996). Besides these two extremes, hedges (*maybe, probably*), and connectors (*but, if*) are used for partial alignment or qualifying and softening stances.

Importantly, keywords indicative of alignment are found to be more abundant than those indicative of disalignment (see Table 8.1), and fewer than 6% of the keyword *agree* collocates with *n't* or *not* to form disalignment, suggesting that expressions of agreement might be more common when users respond to others. Additionally, as shown in the keyword analysis between first contributions and subsequent contributions, the keyword *agree* is more frequent in the first contributions, suggesting that alignment is typically expressed when a user first replies in a thread. However, it should be noted that expression of agreement does not necessarily mean that users agree with the stance. It can be used as a mitigation strategy or a way to indicate relevance or coherence in the discussions (Baym, 1996).

The stance-taking in replies differs from that found in the initiating posts and independent posts because the stance is directed towards specific users who have commented in the thread, instead of course content or general audience. The (dis)alignment with other users' comments indicate users' positioning towards each other via their stance, thus creating user-user interactions that can potentially lead to intersubjectivity (Baym, 1996; Bolander, 2012; Du Bois, 2007; Keisanen, 2007; Kleinke, 2010; Lapadat, 2007).

8.3.2 Meta-language on others' comments

Meta-language keywords found in the replies are mainly those used to refer to others' comments, including *reply*, *posting*, *comment*, *post*, *point*, and the collocation of *you* with communicative verbs *say* and *said*. These meta-linguistic expressions serve three functions. Firstly, users refer to other comments in their replies to maintain coherence within a thread that is polylogal or has evolved into multiple sub-topics, despite the constraints of the threading system (Herring, 1999; Thomas, 2002). This also indicates their engagement with others' comments, as evidenced by the collocation of *thank/thanks/your/you ... reply/posting/comment/post/point*. The keyword *reply* has also been shown in the quantitative analysis to be used significantly more often in subsequent contributions in a thread, suggesting users come back to acknowledge others' replies to them. Secondly, users

employ meta-language to comment on others' language use and commenting behaviours, thus engaging in metapragmatic discourse regarding posting norms in online discussion, instead of discussing the actual content or topic (Tanskanen, 2007). Thirdly, meta-linguistic expressions are also used to point out similarities or differences in viewpoints, thus creating common ground and avoiding communication breakdown (Janier & Reed, 2017; Liu & Liu, 2017; Swales, 2001).

Among these meta-language keywords, there is one unexpected keyword, *link*. Concordance reading shows that it is mostly used to refer to the URLs posted by other users, rather than the dictionary meaning of "connection". An in-depth analysis of URL-posting based on the keyword *link* is conducted in Chapter 9, following Wagner & Herbel-Eisenmann's (2008) approach that focuses on a specific keyword, that is *just* in their corpus, as reviewed in Chapter 3.

In short, the meta-language keywords found in the replies mainly refer to what has been posted by other users, in contrast to the meta-language keywords found in initiating posts (*question, article*) and independent posts (*information, course, knowledge*) which are mainly used for referring to course content. Therefore, the meta-language used in the replies seems to function to establish a dialogic relationship with others who have posted before, thus engaging in user-user interactions.

8.3.3 Interactive language

The interactive nature of the replies within threads is evidenced by keywords which have been established by previous research as used more often in oral language, including *probably, exactly, absolutely*, interjection *ah, oh*, polite speech-act formulae *yes, thanks, thank, hi, sorry, good luck*, or addressing other users with *you* (Biber et al., 1999; Carter & McCarthy, 2006). As revealed in section 8.2, *ah, oh, thank, thanks, yes, you, sorry, reply* are used significantly more often in the subsequent contributions. These keywords may be used when users come back to the thread to respond to others who have addressed them, possibly expressing acknowledgement to others with *reply*, or appreciations with *thank* and *thanks*. This responsiveness towards others who have commented before further establishes dialogue with previous utterances (Bakhtin, 1981).

8.4 Conclusions regarding keyword analysis and the potential importance of disagreement

The functional grouping of the reply keywords identifies three broad discourse practices of user-user interactions in the replies – stance-taking, meta-language, interactive language – all of which are used to establish a dialogic relationship with other users who have commented before in a thread. Unlike initiating posts and independent posts, stance-taking in replies is directed towards other users rather than the course content. Replies contain more meta-language on others' comments compared to both initiating and independent posts that contain more meta-language on course content. In terms of their dialogic nature, replies contain more interactive language which is typically used to respond to previous comments in the threads, whereas independent posts are used to respond mainly to course content, and initiating posts anticipate future potential utterances. This finding thus addresses RQ1 regarding the difference between replies and the two types of posts.

Importantly, the analysis points to the potential role of disagreement and disalignment for stance-taking in sustaining user-user interactions and developing dialogic conversations in FutureLearn, despite the fact that it occurs less frequently than agreement and alignment. The keyword analysis shows that most stance-taking in replies involves expressions of agreement and alignment, as shown by *agree, agreed, yes, true, right* and there are very few negations of these keywords. This finding corresponds to other findings on online discussions in learning settings where users tend to use agreement as a strategy to stay coherent in the threads or to maintain interpersonal relationships with others, rather than challenging each other or engaging in critical discourse (Kellogg et al., 2014; S. Knight & Littleton, 2016; Lapadat, 2007; Paulus, 2006; Rourke & Kanuka, 2007). Still, it is not uncommon for conflict or disagreement to occur in online discussions, as found in other settings, such as commenting spaces on YouTube and news websites or public discussion boards (Bou-Franch & Garcés-Conejos Blitvich, 2014; Kleinke, 2010; Langlotz & Locher, 2012). Indeed, in the current study, although disagreement might be less common than agreement,

the negation and hedging strategy indicated by some of the reply keywords attests to the existence of disagreement in the FutureLearn online discussion.

However, in the present study, keywords indicative of agreement are found more often in replies which are users' first contributions in a thread, compared to users' subsequent contributions in the same thread. This suggests that users are less likely to express agreement when engaging in continued interactions with others within the same thread. Keywords indicative of agreement, *agree*, *yes*, *same*, are also found to be used more frequently in shorter threads than longer ones, suggesting agreement is less likely to lead to sustained user-user interactions in a thread. Similarly, *agree*, *yes* and exclamation mark (!) are used more often in the first reply of one-reply threads compared to those in threads receiving more than one reply. On the other hand, keywords indicative of concession and possibly disagreement, such as *sorry* and *but* are found to be significantly used more often in subsequent contributions. Subsequent contributions happen when users come back to the threads that they have contributed before, thus increasing the chance of them engaging in sustained negotiation with each other. Given these quantitative findings, it can be speculated that agreement, although creating a supportive space and helping to maintain interpersonal relationships among users, might not be conducive for sustained negotiations among users. Three threads where only agreements are expressed will be examined in section 8.5 to provide evidence for this speculation.

The micro-analysis of selected threads in the following section therefore explores the possibility that disagreement can sustain negotiations. As argued in Chapter 3, disagreement can be part of a negotiation process that could potentially move into exploratory talk, intersubjectivity or online deliberation that exposes users to multiple voices and reasoned arguments, depending on users' discourse practices in the negotiation process (Chiu, 2008; Felton et al., 2015; Lapadat, 2007; Lewiński, 2013; Mercer, 2004). As we shall see, disagreement in FutureLearn similarly has the potential of generating dialogic conversations, which are shaped by users' discourse practices in their replies towards others.

8.5 Micro-analysis of threads

Concordance reading of reply keywords is decontextualized in the sense that replies are always in response to comments posted before them and may trigger other replies posted after, thus the functions of reply keywords must be interpreted in light of the whole thread. To contextualize the reply keywords, a series of micro-analyses of threads are conducted in the following subsections, with attention paid to the use of the keywords. Motivated by the findings in preceding sections and extensive reading of threads in the corpus (see section 4.7.3 for micro-analysis procedure), three threads in which users disalign but engage in sustained negotiation with each other are chosen for in-depth micro-analysis. This choice of threads will unravel discourse strategies underlying processes of intersubjectivity as realized by the reply keywords indicative of stance-taking, meta-language and interactive language. Thereby, RQ2 regarding how these discourse practices are employed to sustain dialogic conversations is addressed.

Before presenting the micro-analysis of sustained threads where disagreement occurs, another three threads in which users align with others' stance are first presented to illustrate the agreement practice commonly found in the Futurelearn. As revealed by the keyword analysis, users typically do not come back to engage in sustained interactions after expressing agreement. This contrasts with threads where disagreement arises and users engage in continued discussions.

8.5.1 "I agree"

Two short threads with all replies expressing agreement are compared to show that users sometimes merely express agreement (Figure 8.1), and sometimes contribute their accounts or reasons on top of expressing agreement (Figure 8.2). These two ways of expressing agreement have also been found in independent posts, as discussed in Chapter 7.

Figure 8.1 Expression of agreement in replies.

Thread 4509705

Source: <https://www.futurelearn.com/courses/dyslexia/1/comments/4509705>

Initiating post 2015-04-20 06:45:52 Like: 5

User d1-3123

It's interesting that the issues Dyslexic students may extend into aural aspects of language as well -- new information for me. I like the idea of focusing on receptive subskills and easier tasks, as I feel these emphasize what students CAN understand and process in a text rather than focusing on what they can't, and they help students build an understanding in a process oriented way. I think we often make the mistake of focusing on product over process in learning, and it seems to me that for Dyslexic students, a focus on process is absolutely essential.

Reply 1 (first contribution) 2015-04-20 06:53:14 Like: 0

User d1-2682

Yes - I'm guilty of this.

Reply 2 (first contribution) 2015-04-20 07:34:53 Like: 0

User d1-4216

I would strongly **agree** with that!

Note. Emphasis is mine, the same goes for other figures in this chapter.

Figure 8.2 Expression of agreement in replies with additional content.

Thread 17120574

Source: <https://www.futurelearn.com/courses/palliative/1/comments/17120574>

Initiating post 2016-10-27 11:37:16 Like: 3

User p1-93

I suppose that in many ways Palliative and EoL are very much seen as the same thing in the Care home where I work. Simply because we do not understand the difference in many ways. We the carers have not really received any training on this. I noticed that through the course so far the education being discussed is all for Dr or Nurses, I would have thought it was important for many other people to get some training as well. Social workers especially so that they are aware what is available and out there when they are assessing their clients, care workers so that we understand our place in this strategy and are given the tools to talk to residents and family members about end of life decisions.

I think also that there needs to be some over all information given to the general public (just don't ask me how) I am sure that at some point this becomes an issue in all our lives (be it for us, a relative or friend) and that emotional time is not the best time to get your first introduction to the subject.

Reply 1 (first contribution) 2016-10-27 16:49:41 Like: 1

User p1-336

I *agree* [p1-93] that public knowledge is very necessary. Primarily so people will know what's available and give political input for what is not. Information for the public can be by publications in general magazines, documentaries and discussions on television, pamphlets in medical offices and facilities. Those involved in PC should push to have such information given to the public.

Reply 2 (first contribution) 2016-10-27 18:05:45 Like: 1

User p1-85

I *agree* as well. Many people in the area where I live think that Palliative Care means "placebo", something that doesn't work, and they are terrified of going to a hospice because it means they "are about to die"... I find myself having to explain to patients and families what all this is about.

Threads from Figure 8.1 and Figure 8.2 both are from steps with a discussion prompt. The initiating posts of both threads can be deemed as responding to the prompt, whereas the replies in each thread can be seen as aligning with the initiating posts on the subject. In Figure 8.1, the replies are mere expression of agreement without any additional content, "Yes - I'm guilty of this.", "I would strongly *agree* with that!". Mere expression of agreement of similar patterns have been found in

319¹³³ instances out of the 4373 instances of *agree* or *agreed* in all users' replies. Despite not including additional content, 76 of the 319 replies gains at least one likes, of which one gains five likes, and another gains four likes.

In contrast, in Figure 8.2, each reply starts with an expression of agreement, "*I agree...*", followed by additional contents. In reply 1, the user p1-336 further answers the question raised in the initiating post, "just don't ask me how", whereas in reply 2 the user p1-85 shares personal experience, "I find myself having to explain", to support their stance and alignment with what has been said in the thread. Despite this difference between these two threads, neither thread develops further into sustained threads. Besides short threads, a few long threads also contain mainly expression of agreement, but show no evidence of negotiation, as illustrated in Figure 8.3.

¹³³ The count is based on the regular expression `<post>[{}1,5]{word="agreed|agree"%cd}[{}1,5]</post>`, which search for replies with the keywords *agree* and *agreed* with fewer than 5 words from the start and the end of the replies. This is based on an arbitrary assumption that sentences fewer than 10 words would not have much elaboration.

Figure 8.3 Expression of agreement in replies in a long thread.

Thread 18861825

Source: <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/18861825>

Initiating post 2017-01-16 03:48:25 Like: 39

User n4-1154

I don't consider obesity a disease but it is a complicated imbalance of factors involving hormones, genetics, gut bacteria, emotional, stressful or psychiatric issues, effects from medications, environmental factors, in conjunction with a mechanized and less labor-intensive "evolved" modern existence and the changes in our food supply (processed foods, fast foods, eating out and agricultural changes which have altered the nutritional content of our food).

Reply 1 (first contribution) 2017-01-16 04:02:04 Like: 3

User n4-2234

Totally agree with *you*!

Reply 2 (first contribution) 2017-01-16 04:05:55 Like: 2

User n4-1566

I *agree*.

Reply 3 (first contribution) 2017-01-16 04:28:33 Like: 2

User n4-491

I think this puts it very well

Reply 4 (first contribution) 2017-01-16 06:33:09 Like: 0

User n4-1934

Obesity is the modern disease leading to many other diseases.

Reply 5 (first contribution) 2017-01-16 08:09:56 Like: 3

User n4-3128

Yes agree with this. Wouldn't class as disease. I also believe genetics plays a big part in it *too*

Reply 6 (first contribution) 2017-01-16 08:19:26 Like: 0

User n4-1133

For many years I've known a woman who is obese and she has applied all methods to lose weight but she couldn't. I advised her to go on a veggie diet (which I myself experienced helping me lose 4 kg within 3 months) but she said she couldn't do that, though she tried but if she had eaten less she could be hospitalized! Now she is both obese and suffering from diabetes. What is the right solution for her?

Reply 7 (first contribution) 2017-01-16 21:38:16 Like: 1

User n4-1603

A woman I know had surgery and she is very glad she did. This can be considered as a first step and should be followed by a healthier lifestyle regarding exercise and nutrition.

Another approach would be fecal microbiota transplant. Obese people have a distinct imbalance in their gut microbiome. Through manipulation of the bacteria that live in the gut we can have an impact on obesity. This is not mainstream yet, but I think it probably will be in a few years.

Reply 8 (first contribution) 2017-01-18 21:49:06 Like: 4

User n4-3021

[n4-1154]

I *agree* with all the factors you list as being influential in causing obesity, *but* a large number of these are beyond our control, and as such I would class obesity as a disease.

Reply 9 (first contribution) 2017-01-25 20:20:16 Like: 0

User n4-235

I *agree*.

The thread in Figure 8.3 occurs in a step with a discussion prompt: “Do you think that obesity is a ‘disease’? What do you think is meant by the term ‘disease’? [.....] Do you think that obesity is a mis-match between our physiology and environment?” The initiating post explicitly responds to the prompt with their stance, “I don’t consider obesity a disease”. The first three replies are simply agreement, “*Totally agree* with you!”, “I *agree*”, aligning with the initiating post, although reply 3 does not use the keyword *agree*, “I think this puts it very well”. Reply 4 and 8, although without disagreement tokens or negations, contain statements “Obesity is the modern disease”, “I would class obesity as a disease” that contradict the initiating post and reply 1 to 3. Similarly, reply 6 and 7, based on the examples shared, seems to be against the initiating post, although there is no explicit mention of disagreement or agreement. It is ambiguous which stance reply 9 aligns with since no specification is given after “I *agree*”. This thread also attests to the difficulty of determining to whom each user replies in a polylogue when users do not employ discourse practices such as meta-language to refer to another comment or indicate their addressee.

Although there are two stances raised, obesity is vs. is not a disease, no negotiation is generated in this thread. Also, although reply 7 and 8 seem to raise examples for exploration of this issue, they do not trigger further discussion. Within the whole thread, nobody comes back again to the thread, which is commonly the case according to the quantitative description of the corpus in Chapter 5. In short, although disagreement and examples are raised among voices of agreement, no negotiation happens between the two stances. This might be because nobody comes back to engage in different viewpoints such that no stance is highlighted, especially when the disagreement is not explicitly stated, and opposing stances are not questioned. To some extent, although expressions of agreement indicate users responding to each other, this conversation borders on parallel monologues since nobody takes up others’ stances or examples, despite there are nine replies, and the initiating post receives 39 likes.

The brief analysis of these three threads shows that, although expression of agreement creates an interactive and supportive space, sustained negotiation and intersubjectivity might not

occur. This observation is consistent with other findings on online discussions in learning settings where users tend to express agreement rather than challenge (Kellogg et al., 2014; Lapadat, 2007; Littleton & Whitelock, 2005; Paulus, 2006; Rourke & Kanuka, 2007). Mere expression of agreement, according to Mercer (2004), is cumulative talk that is not conducive for dialogic learning and exploration of alternative voices. At the same time, the analysis of agreement thus far seems to differ from Baym's (1996) finding that agreement can be a way of mitigation before raising one's opposing stance, probably users in MOOCs seldom use this strategy to raise their disagreement, similar to the other online learning spaces mentioned earlier.

Therefore, in the following subsections, more in-depth analysis of three threads where disagreement is explicitly raised are conducted to examine how negotiation and intersubjectivity can happen in disagreement, although the instances of disagreement are rarer than agreement based on keyword analysis. Each of the micro-analysis starts with an overview of the context and participation patterns of the users in the thread, the occurrences of reply keywords in the thread, and the start of the thread (the initiating post and first reply). Then the discourse practices found in the threads are discussed in relation to the keywords found. How each thread ends is also discussed. Differences and similarities in discourse practices underlying these three threads are then drawn.

8.5.2 “True but...” Concession strategy in negotiation

8.5.2.1 Context

The first thread¹³⁴ to be examined is the longest user-only thread in the contract-4 course, in a step without a discussion prompt but a video where a course educator introduces the law of contracts and intellectual properties¹³⁵. The thread contains 11 replies involving mainly two users who repeatedly come back to the same thread: the initiator who contributed one initiating post and five replies, and another user five replies. There is also one reply from another user, which happens towards the end of the thread (Reply 9), who might act as a mediator because after this user's reply,

¹³⁴ <https://www.futurelearn.com/courses/contract-management/4/comments/18016862>

¹³⁵ <https://www.futurelearn.com/courses/contract-management/4/steps/115131>

the two repeated users seem to come to a reconciliation (Reply 10 and 11). It could be argued that this thread is mainly a dialogue between the two users who keep coming back to the thread. In this thread, they engage in turn-taking, i.e., responding every time after another replies, either on the same day or the next day¹³⁶.

8.5.2.2 Keywords

Two keywords *but* (n=8) and *same* (n=7) are used frequently by both users in their discussion. *But* is always used in a concession practice in their negotiation with each other where they re-instate their own voice after acknowledging others' voice with attitude keywords such as *oh*, *yes* and *true*. *Exactly* and *too* are also used by users to emphasise their own stance. *Same* is used in relation to the topic of discussion, whether a fake product is the same as the original product.

8.5.2.3 Start of the thread

The initiating post and the first reply are posted by the same user m4-285 (Figure 8.4). Interestingly, the initiator seems to answer their¹³⁷ own question in the first reply.

Figure 8.4 Start of thread 18016862.

Initiating post 2016-11-21 Like: 3

User m4-285

About this: [...] Trademarks need registration. But that registration has an indefinite duration. And a trademark is simply some way to represent the entity to which it is associated. These can be in words, in the form of logos, shapes and packaging, smells, sounds, actions or slogans. [...] I had no idea that a «smell» *could* be a trademark... (sound I am guessing like some of the iPhone rings or notification sounds for *example*... But smell? Does *anyone* have an *example*? And what about taste then?

Reply 1 (initiator's subsequent contribution) 2016-11-21 Like: 0

User m4-285

Ohhhh! A perfume!! Like Chanel #5 ... Ok nevermind that! So what about taste?

¹³⁶ The provider of this course did not provide the exact timestamp but only the date of each post, so it is hard to tell whether the two users engage in an almost synchronous interaction or asynchronous interaction.

¹³⁷ Their is used for his/her given that the user's identity is anonymized as well as their gender.

In the initiating post, the initiator m4-285 first quotes verbatim a statement from the course video, “Trademarks need registration..... smells, sounds, actions or slogans”, creating a common ground for establishing a dialogue with a potential audience, especially in a step without any discussion prompt. Following this common ground, the initiator’s “I had no idea...” shows an uncertainty in their epistemic status to establish a dialogic relationship with others who have a knowing status. But this uncertainty status seems to serve as taking a stance against “smell could be a trademark”, rather than a genuine question. The initiator contrasts this uncertainty to a more certain example “sound I am guessing like some of the iPhone rings...”. The question “But smell?” that follows could be a rhetorical question that states a stance (Kleinke, 2012). Lastly, before ending the initiating post, the initiator’s question “Does *anyone* have an *example*?” seems to open up the space for any of the users to pitch in with any suggestion, with initiating keywords *anyone*, *example* and question mark (?). The discourse practices of showing uncertain epistemic status, addressing potential users with *anyone* and inviting *example* have been discussed in Chapter 6 as discourse practices that could possibly increase the chance of receiving replies and initiating a conversation.

8.5.2.4 Discourse practices in replies

In reply 2 to reply 6 (Figure 8.5), initiator m4-285 and user m4-153 engage with each other’s point of view through a discourse practice of conceding and reasserting. They concede by first agreeing with the other’s viewpoint, then reasserting their own view with the reply keyword *but*, and challenging the other’s view.

Figure 8.5 Reply 2 to reply 6 of thread 18016862 where concession practices are observed.

Reply 2 (first contribution) 2016-11-22 Like: 1

User m4-153

Cocacola?

Reply 3 (initiator's subsequent contribution) 2016-11-22 Like: 0

User m4-285

Mmmmm Ok... **But** then who decides, in these cases, that the taste is **too** similar to Coca Cola of that this fake perfume is **too** similar to the original one... There is generic Cola... There are perfume copies... I mean, when it comes to smell and taste **isn't** it partly subjective? I have heard before cases where there is litigation for a song that is **said** to have been copied and that violated copyrights and then they send an expert on harmonics or something to evaluate the level of similarity with the «original»... **But** taste and smell... ??

Reply 4 (subsequent contribution) 2016-11-22 Like: 2

User m4-153

Yes, there are copies, **but** they cannot be the **same** as the real thing. A very good imitation LV handbags, no matter how good it was copied can never be the **same** as the real thing. Each LV bag is hand made and the monograms are arranged in a unique way. even though the imitated ones are made of quality leathers, such little details give it away. Also for taste, there are thousands of other cola drinks, **but** they are never the **same**. Channel No 5 might be imitated, **but** it wouldn't smell the **same** as. also the longevity of the smell might differ **too**.

Reply 5 (initiator's subsequent contribution) 2016-11-22 Like: 0

User m4-285

That is **true but** that is not **exactly** what I was wondering... Is taste really a trademark **if** you can replicate it? Coca cola does not go behind generic colas and sue them for violating their property. Nor do the perfume companies... Do they have people specialized in evaluating the other colas and **say: oh** ok, it tastes different enough from Coca Cola... ? **Same** for perfume...

Reply 6 (subsequent contribution) 2016-11-22 Like: 0

User m4-153

I understand *your* stance. They can't be sued because they **don't** taste or smell the **same**. Close enough, **but** not the **same**.

The first concession is found in reply 3 when the initiator m4-285 takes up the m4-153's response "Cocacola?" in reply 2 to their question in reply 1 "So what about taste?". The concession is indicated by "Mmmmm Ok" in which *Ok* could mean agreement, yet the following "But then who decides, in these cases, that the taste is *too* similar to Coca Cola" indicates initiator m4-285's reassertion and disagreement with m4-153's response. Following this concession and reassertion,

the initiator m4-285 also further elaborates on their stance by pointing out “generic Cola” in response to user m4-153’s “Cocacola?”, raising a rhetorical question “when it comes to smell and taste isn’t it partly subjective?”, and presenting a contrasting example of song. It could therefore be argued that “Mmmmm *Ok but...*” is a concession strategy that is used to engage with the previous utterance and other readers before expressing disagreement and reasserting one’s own stance.

Similarly, in response to the initiator, user m4-153 also makes a concession “*Yes...but...*” before launching into a series of examples to reassert their claim that copies are “never the *same*”. User m4-153 uses the pattern negation + “*same*” four times: “cannot be the *same*”, “never be the *same*”, “never the *same*”, “wouldn’t smell the *same*”. This repetition and parallelism can be a rhetorical strategy of emphasising one’s own stance and making evaluation of others’ stance (Tannen, 2007). Besides raising a new example “LV bag”, user m4-153 also refers back to the example “Chanel No 5” given by the initiator and to “cola” that has been discussed in the thread, this repetition and intertextuality creates coherence in the thread and indicates that the other’s points are taken up, thus maintaining their dialogic relationship in the conversation. Again, in response, the initiator m4-285 first uses a concession to engage with user m4-153’s response “That is *true but...*” before repeating their initial question in the initiating post but with a conditional to refine it to be more specific, “is taste really a trademark *if you can replicate it?*”. In turn, user m4-153 responds with another concession strategy, “I understand *your stance*” yet again restates their own claim with *but*, *not* and *n’t*. “*Your stance*” is a meta-language similar to *your point*, which is the reply keywords.

This “*yes...but*” practice is similar to the other-trigger concession found in oral conversations (Lindström & Londen, 2013). Other-trigger concession is a response and acknowledgment to the other’s opposing views, thus suggesting the dialogic nature of the concession practice in this thread because voices of others are also referred to in one’s replies. This other-trigger concession can also be seen as a way of staying coherent in the discussion while users pitch in with more supporting examples and disagree with each other, although they never explicitly mention they do not agree

(Pomerantz, 1984). Prefacing one's reply with agreement to each other may also maintain an interpersonal relationship with others (Baym, 1996). All these point to the possibility of intersubjectivity in the negotiation process where each user considers other's subjectivity while tabling theirs, and the reciprocity and interactive nature of concession. Therefore, although Antaki & Wetherell (1999) argue that concession can be just a "show", this thread shows that the concession allows both users to elaborate on their stance in relation to the other's view and creates a space that allows multiple voices, supporting Lindström & Londen's (2013) argument that concession is an important strategy for reasoning and argumentation.

8.5.2.5 End of the thread

The turn-taking of concession and reassertion between the initiator m4-285 and user m4-153 stops in reply 7 when the initiator m4-285 asks a direct question and another user m4-443 joins the discussion which seems to bring the two users into alignment (Figure 8.6).

Figure 8.6 End of thread 18016862

Reply 7 (initiator's subsequent contribution) 2016-11-22 Like: 0

User m4-285

Who draws that line?

Reply 8 (subsequent contribution) 2016-11-22 Like: 0

User m4-153

The Patent, TM or IP details does. I suppose.

Reply 9 (first contribution) 2016-11-23 Like: 1

User m4-443

I was wondering about this - must be very difficult to draw an **agreed** line in cases where there is an area of subjectivity.

Reply 10 (subsequent contribution) 2016-11-24 Like: 1

User m4-153

True. But Subjectivity is Individually Contrued

Reply 11 (initiator's subsequent contribution) 2016-11-24 Like: 0

User m4-285

Exactly!

In reply 7, the initiator m4-285 asks a short “who” question which requires an exact answer. This question, in fact, is similar to the one that has been embedded in their long reply 3 to user m4-153, “*But* then who decides...the taste is *too* similar to Coca cola...”. In this short turn in reply 7, the initiator repeats the core question within their initiating post, but does so baldly, without any concession. This short question is different from the previous use of yes/no interrogative questions, “Isn’t it..”, “is taste really...”, “Do they have...”, or open-ended questions, “But smell?”, “What about taste?”, “But taste and smell...?”, and can be considered as serving as a challenge while maintaining one’s own stance. User m4-153 replies dutifully to this “who” question with a specific answer, though ends it with a hedge “I suppose”.

At this point, after one day, another user m4-443 joins this conversation. This participation shows that, although this conversation has been exclusively contributed by the initiator m4-285 and user m4-153, the other users have been reading it. This is further suggested by the new joiner’s cataphoric reference to the two concepts mentioned earlier in the thread: “(agreed) line” in reply 7 and “subjectivity” in reply 3. In this case, user m4-153 agrees with a concession “*True. But...*” before making the claim that “subjectivity is individually Contrued”, which the initiator then aligns with. It seems that the initiator m4-285 and user m4-153 have converged to the same understanding towards the end of the thread.

8.5.3 “I did not say...” Meta-language and *if*-conditionals

8.5.3.1 Context

This thread¹³⁸ occurs in a step in nutrition-4 course with a discussion prompt: “Health shops and some magazines often heavily promote the use of protein and amino acid supplements for building up muscles. Do you think there is merit in these practices? Does more (protein) mean bigger (muscles)?”¹³⁹ The thread contains seven replies involving mainly two users who repeatedly come back to the same thread: the initiator who contributes the initiating post and two replies, and

¹³⁸ <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/19605035>

¹³⁹ <https://www.futurelearn.com/courses/nutrition-wellbeing/4/steps/124762/>

another user four replies, one of which is directed to another user who contributes once in the thread. Similar to the previous thread, this thread starts from a dialogue between two users and moves towards a polylogue.

8.5.3.2 Keywords

The pronoun keywords *you* (n=16) and *your* (n=6) are frequently used in this thread. Five of the pronouns are used by users to address each other, especially when used with meta-language keyword *point* (n=2), communicative verb keywords *say* (n=3) and *said* (n=2), other forms which are not keywords, *saying* (n=2) and *says* (n=1). These are mainly used in commenting what one or other has said before in their negotiation of understanding each other's view. Besides, there are nine *ifs* used in the thread, all of which are used as conditionals to qualify arguments. Lastly, three of the six *buts* found are used in concession strategy described earlier.

8.5.3.3 Start of the thread

The initiating post and the first reply are presented first (Figure 8.7) to illustrate the disagreement between the two users over the consumption of protein before the discourse practices of their negotiation in replies are introduced.

Figure 8.7 Start of thread 19605035.

Initiating post 2017-02-12 08:48:19 Likes: 1

User n4-1568

If you get the maximum amount of protein through healthy eating then supplements with added protein are clearly **just** a money making scheme as you flush them out right away. I did not know this so interesting to hear

Reply 1 (first contribution) 2017-02-13 18:31:15 Likes: 1

User n4-1472

Some people feel the **same** way, many people on here have commented almost **exactly** that. What I have found, from working in a gym, selling supplements, and taking them, is that it depends on the individual, and their goals.

After all, *if* a balanced diet was sufficient - we would not have the obesity rates that we do?

If everyone had time for balanced diets, we wouldn't have the obesity rates?

If people could afford these healthy, raw, unprocessed, organic foods - we wouldn't have obesity related health issues?

Introducing quality proteins, fibre, correct sugars, vitamins and minerals (that your body can not make) is a great way to educate people about diet and nutrition. I think it gets results, and people can see the benefits. Then the challenge is to replace the supplements with real foods - again depending on the person's goals...

Does the average person have the time, money and inclination to get the **right** quantity, quality of all macro and micro nutrients in their daily diets?

The initiating post seems to be a response towards the discussion prompt with an *if*-conditional to qualify their rather strong stance “clearly just a money making scheme”. The post also ends with an indication of their “becoming knowing” status “I did not know this so interesting to hear”. To some extent, the discourse practice in this initiating post is more similar to the independent post, as concluded in Chapter 7. However, it attracts objection from user n4-1472. It is possible that it is the content of this initiating post that increases its chance of receiving replies, or probably because of this replying user. It is found that the replying user n4-1472 has been posting the same first reply in response to nine other posts, within a time span of six minutes in this step. None of these replies receives response from the initiators of those posts or other users such that those threads end at this user’s reply except this thread under examination. This difference suggests

the importance of users coming back to engage with others, as shown in the subsequent contributions by both the initiator and this replying user in this thread, as analysed next. However, in this reply, user n4-1472 does not explicitly mention their disagreement towards the initiating post but a qualifying argument “depends on the individual” and possibly the rhetorical question at the end of the reply 1 “Does the average person have the time [.....]to get the *right* [.....] diets?” can serve as a stance-taking (Kleinke, 2012).

8.5.3.4 Discourse practices in replies

Two discourse practices, qualification with *if*-conditionals and meta-language are found in the negotiation process between the initiator n4-1568 and user n4-1472 in reply 1 to reply 5 (reply 1 is in Figure 8.7, whereas the other replies are in Figure 8.8).

Figure 8.8 Reply 2 to reply 5 of thread 19605035 where qualification and meta-language are observed.

Reply 2 (initiator's subsequent contribution) 2017-02-13 19:15:52 Likes: 0

User n4-1568

I see **your point but** I was also thinking about these people that work out a lot eat a lot of protein and then get the supplements to go with it. **If** people **don't** get the protein in through eating of course a supplement could be beneficial, **but if you** already eat protein with every meal then it's **just** a waste of money really **don't you** think?

Reply 3 (subsequent contribution) 2017-02-13 19:58:28 Likes: 0

User n4-1472

It does entirely depend on **your** goals really. For the people in the gym looking to 'bulk up' and I was one of them, and **you** are exercising a lot, **your** muscles are working harder, and they need to repair in order to grow. To get the amount of calories needed, and the amount of protein, as discussed in the previous sections - takes a lot of eating! *Maybe* 4 or 5 chicken breasts a day. **If you** can get that in a well balanced shake - it is easier to consume, cheaper, **if** bought in bulk and uptake is quicker (I think) - and **you** have the added bonus of drinking the shake after the work out on the way to work, or home. Getting home, or to work, then cooking is not always ideal.

With most of these products it is about convenience - ideally we should fuel the body throughout the day - little and often - 5 or 6 meals a day - so a shake or a bar that counts as one of these meals, again, is ideal (convenient) - also stops **you** snacking on things like crisps or chocolate at work! There is always a birthday in our office with cake and biscuits :) There is a place for it I think - and to **say** it is a waste, is to **say** thousands are wrong...

Reply 4 (initiator's subsequent contribution) 2017-02-13 20:21:50 Likes: 0

User n4-1568

I did not **say** it is a waste **but** it seems to be a waste **if you** dispose of it without taking any benefits from it. Just like overdosing on vitamin c would be a waste as **your** body **doesn't** use more than a certain amount and disposes of the rest. Neither am I saying so many people are wrong, I used to drink protein shakes myself when I was not eating many proteins. All I was saying that according to the video we need x amount and anything taken extra **doesn't** get used, therefore **if** a person gets sufficient protein throughout normal food, adding protein bars or shakes to your diet sounds like a waste of money to me. Without a doubt **you** know a lot more about it than I do, it was a simple connection I made between what was **said** on the video and what I have seen people around me do.

Reply 5 (subsequent contribution) 2017-02-13 21:15:49 Likes: 1

User n4-1472

sorry, I thought you **said** supplements with added protein are clearly a money making scheme. It does make **you** think though, which is the point of the course :)

We had a protein shake with 53g of protein in it, **but** we made sure people only had half the bottle, and refrigerated the rest before bed - because ultimately **you** are **right**, and the video says it - **you** can overdose, and **you** can only synthesise so much...

If-conditional is used to create alternative scenarios to support their arguments for or against protein supplement in the first three replies. In reply 1, user n4-1472 uses a repetition of three rhetorical questions with a structure of “*if....., we would not/n’t.....?*”. Initiator n4-1568 also uses a similar strategy in reply 2 with two *ifs* in a sentence to specify different scenarios, and ends it with a question “*If people don't get the protein in through eating of course a supplement could be beneficial, but if you already eat protein with every meal then it's just a waste of money really don't you think?*”. User n4-1472 further uses the *if-conditional* to specify an argument in reply 3 “*If you can get that in a well balanced shake - it is easier to consume, cheaper, if bought in bulk and uptake is quicker*”. In short, the initiating post and the three replies are built on *if-conditionals* when users are debating with each other while limiting their arguments to a certain condition. Importantly, the qualification of one’s own argument with *if-conditionals*, rather than bare assertions, seems to help their resolution in the reply 4 and 5 when they clarify their (mis)understanding of each other’s argument, as described below.

The meta-language, realized by the keywords *you, say, said* and non-keyword *saying* are used when users clarify what they have said before. The clarification starts after user n4-1472 comments in reply 3 “*to say it is a waste, is to say thousands are wrong...*”, which seems to be interpreted by the initiator n4-1568 as an accusation to them. This is evidenced by the initiator’s clarification in reply 4 in response to this accusation where the initiator repeats what has been said by a series of ‘say’ construction, “*I did not say*”, “*Neither am I saying*”, “*All I was saying*”, and also qualifies what has been said with two *if-conditionals* “*but it seems to be a waste if you*”, “*if a person gets sufficient protein throughout normal food.*” This *if-conditionals* specify the limits of their argument, which seems to be interpreted by n4-1472 as sweeping generalization beforehand, as evidenced in reply 5 where user n4-1472 apologizes retrospectively for misunderstanding, and uses another meta-language keyword “*sorry, I thought you said supplements with added protein are clearly a money making scheme.*”

Both users also employ meta-language to comment on their own epistemic status in their clarification to each other. The initiator n4-1568 makes it clear that “it was a simple connection I made between what was *said* on the video and what I have seen *people* around me do.” in reply 4, while n4-1472 seems to suggest that n4-1568’s comment “does make [*them*] think though, which is the *point* of the course :)”. Revealing their own epistemic status to each other can be considered as one way of creating intersubjectivity.

Beside epistemic status, it is also noteworthy that the initiator n4-1568 seems to compromise their stance against protein consumption by resorting to personal experience “I used to drink protein shake”. This can be considered as an identity performance, and one way of intersubjectivity to indicate one’s acceptance of the other’s subjectivity, thus a resolution. Besides this personal experience, user n4-1568 also attributes epistemic authority to user n4-1472 by saying “Without a doubt *you* know a lot more” in reply 4, which seems to be in response to the identity work by user n4-1472 in reply 1 “working in a gym”. In response, user n4-1472 also reciprocates this epistemic attribution by “ultimately *you* are *right*” in reply 5, and refers back to the video as the initiator n4-1568 does in reply 4. This strategy of attributing epistemic authority to others also indicates acceptance of others’ subjectivity, thus creating intersubjectivity and a resolution.

8.5.3.5 End of the thread

There are two more replies after the turn-taking between initiator n4-1568 and user n4-1472, which is contributed by a newly joined user n4-346 and responded to by user n4-1472 (Figure 8.9).

Figure 8.9 End of thread 19605035

Reply 6 (first contribution) 2017-02-14 20:21:48 Likes: 1

User n4-346

even without this supplements there are so many high protein items available and can be found in our daily meal

Reply 7 (subsequent contribution) 2017-02-14 20:42:46 Likes: 0

User n4-1472

Yes. Yes there are.

I **totally agree** with **you**.

But Supplements have a benefit - they can pack everything the body needs into a shake, and when eaten with healthy foods, they can have transformational effects.

Not everyone has the time, or money to get all the macro and micro nutrients into their diets - or the knowledge... How much Mg? Na? protein? Pulses? Wholegrain? Fat - saturated or unsaturated...

Education is an issue - I have seen many people start on shakes and bars, SEE the difference a healthy diet with added nutrition can have. Then we educate, and they find natural alternatives :)

Reply 6 by user n4-346, “many high protein items..... can be found in our daily meal” seems to be more aligned with what initiator n4-1568 has said in reply 2 “If you get the maximum amount of protein through healthy eating then supplements with added protein are clearly *just* a money making scheme”. In response, user n4-1472 first concedes with “Yes. Yes there I *totally agree* with *you*”, then reasserts with “*but.....*” and includes both voices “people start on shakes.....and they find natural alternatives” in this last reply. The participation of the n4-346, as with the previous thread, shows that a dialogue between two users has also been read by other users, although the discussion does not continue further afterward.

8.5.4 “You have never *replied* to my original *point*, but that would make a good academic discussion!” Metapragmatic discussion

8.5.4.1 Context

This thread¹⁴⁰ happens in a step in ancient-1 course with a title “When does a food become a drug?” and contain mainly discussion prompts “Can you think of examples of foods that you use medicinally? What diseases do you use them for? What gives them healing properties? Are these foods always healthy?”¹⁴¹. This thread only consists of six replies contributed by two users, ah1-365 and ah1-993 who engage in turn-taking. User ah1-993 is the third most prolific user, i.e., a super-poster, in the discussion of this MOOC, who contributes 368 comments in the discussion, 98% of them are replies. The initiator of the thread ah1-216 never comes back to this thread, which is quite common in the online discussions, as shown in Chapter 5.

8.5.4.2 Keywords

The most frequent reply keyword found in this thread is *you* (n=20), 15 of which are used by users to address each other. The other keywords used frequently in this thread are *n’t* (n=13) and meta-language keywords *say* (n=5), *reply* (n=1), *comment* (n=1), *post* (n=4), *point* (n=3). There are also non-keywords which are the derivatives of the meta-language keywords, such as *says*, *saying*, *comments*, *posts*. These keywords are used with *you* by the two users when they express disagreement with each other, although to some extent they can be accusatory in nature, similar to disagreement in oral conversations and in some online forums (Janier & Reed, 2017; S. Scott, 2002; Sotillo & Wang-Gempp, 2016). There are nine instances of *if* in this thread, however only one of them is used in a conditional, and the others are used as conjunction to mean “whether”, suggesting there might be lack of qualification and hedging of arguments, compared to previous two threads where *if*-conditionals are used.

¹⁴⁰ <https://www.futurelearn.com/courses/ancient-health/1/comments/20392679>

¹⁴¹ <https://www.futurelearn.com/courses/ancient-health/1/steps/154843>

8.5.4.3 Start of the thread

The initiating post and the first reply are presented first (Figure 8.10) to illustrate the issue that generates the discussion, although the thread develops into metapragmatic discussion afterwards.

Figure 8.10 Start of thread 20392679

Initiating Post 2017-03-08 04:23:05 Likes: 1

User ah1-216

My dad always swore by Cayenne pepper a tomato juice for a sore throat. "If it doesn't kill you, it'll cure you." One of the latest medicine food fads is apple cider vinegar for everything from acid reflux to lowering your blood pressure. I personally avoid anything that comes with 'miracle cure' on the label. My wife has started using it and just scowls at me when I laugh. I just hope she at least some kind of placebo affect from it.

Reply 1 (first contribution) 2017-03-20 18:01:11 Likes: 0

User ah1-365

Apple cider vinegar has been around a long time and *isn't just* a recent fad. I have arthritis in my hands and, trust me, it works it *isn't* a cure as there is no cure for arthritis *but* the difference in the "before" and "now" is immense. Everyone is different *but* I hope it works for *your* wife, whatever, she is using it for.

Cayenne pepper - it was the major ingredient for a detox fad a few years ago. Wonder *if* that worked for those who used it - *can't* think of anything worse

:o)

The initiating post seems to be posted in response to one of the discussion prompts in this step and raises two examples "cayenne pepper" and "apple cider". The initiator ah1-216 takes a stance against apple cider by saying it is one of many 'fads', a 'miracle cure' and 'placebo'. The initiator never comes back to the thread, probably because they are just replying to the prompt rather than intending to engage with others' replies. Nonetheless, these two examples might form a common ground for dialogue, although the initiating keyword, *example*, is not used. One of the examples, "apple cider" indeed becomes the topic of discussion in this thread when user ah1-365 opposes the initiator by replying "*isn't* just a recent fad" and "it works" with personal experience on

its effect on arthritis. Yet, in this reply, user ah1-365 also qualifies this claim with “it *isn't* a cure” “Everyone is different”.

8.5.4.4 Discourse practices in replies

Reply 1 contributed by user ah1-365 is the start of turn-taking between ah1-365 and ah1-993 until the end of this thread (Figure 8.11 shows reply 2 to reply 5). The exchange is mainly meta-language on what one has said, which realizes metapragmatic discussion about what one should write in an “academic” discussion, as the users put it. This meta-discussion is achieved by using the reply keyword *you* to comment on each other’s replies.

Figure 8.11 Reply 2 to reply 5 of thread 20392679 where metapragmatic discussion happens

Reply 2 (first contribution) 2017-03-20 20:52:34 Likes: 1

User ah1-993

The problem is that the symptoms of arthritis wax and wane. Bad periods are often followed by good and **you can't** really know **if** it is the vinegar that is helping or **if** this is **just** a natural fluctuation in the condition. Which type of arthritis might be relevant - **you** shouldn't really recommend something and **say** it helps 'arthritis' without being specific - osteoarthritis or rheumatoid? Two very different diseases.

Reply 3 (subsequent contribution) 2017-03-21 14:05:21 Likes: 1

User ah1-365

[ah1-993],

I know how my arthritis "works". I also know the difference between Osteo and Rheumatoid. I was not recommending [Initiator A]'s wife takes it - she is already taking it - I was **saying** it works for me. I would not deem to ask someone I do not know what her condition is.

Note, I also clearly stated:

"Everyone is different **but** I hope it works for **your** wife, whatever, she is using it for."

Reply 4 (subsequent contribution) 2017-03-21 15:57:52 Likes: 0

User ah1-993

[ah1-365], I wasn't querying what **you said** about [Initiator A]'s wife or what has happened to **you**. The **point** I was trying to make is that when **you** have a condition where the symptoms are known to fluctuate, it is very difficult to **say** whether any particular alteration in **your** lifestyle has had an effect, or whether that improvement would have happened anyway.

In another **post**, about apple cider vinegar, **you** specifically **say** 'Two words - Take It.' In this **post** you tell us 'trust me, it works'. That sounds to me as **if you** are recommending it to anyone who reads that **comment**.

I didn't **say** that **you** didn't know the difference between osteo and rheumatoid arthritis, or that **you** don't know there are other types of arthritis as well, only that **you** made a blanket recommendation for all.

Don't you think that on an academic website we should discuss the pros and cons of using something that can have adverse side effects **if** used in excess?

Reply 5 (subsequent contribution) 2017-03-22 10:06:59 Likes: 1

User ah1-365

[ah1-993],

Ahh, my enthusiasm can get the best of me as I do believe in the benefits of ACV and, **yes**, I did **say**, "Take It" as I'm a firm believer in its' benefits.

I believe that people will use their common sense when reading a post on any forum and would research a product first before taking it. **Maybe** I should not give people credit to not blindly take something that a stranger has enthused about.

I detest fish so to benefit from their oils take supplements **but** I did not walk into H&B and **just** buy the first thing without looking into whether it was safe for me to do so.

Yes, I **agree**, academic websites are all about discussions and debates. It seems as though **you** consider my phrasing to have been flippant **but you** don't have to **post** long, in-depth comments on everything, my enthusiasm took over and, perhaps, my trust in people to be sensible was misplaced.

[ah1-365]

PS: Note I also **said** "Everyone is different....." that, to me, **says** a person will think "might be worth looking into to see **if** it may help me" and I know very few people who go on a site and don't follow other links about that product and each site **you** go on **you** find out more and peoples' opinions albeit good or bad.

Besides reply 1 contributed by ah1-365, the only discussion on the apple cider within the exchange between user ah1-365 and user ah1-993 is in reply 2 where user ah1-993 opposes ah1-365's stance by specifically zoning in on arthritis "The problem is that the symptoms of arthritis wax and wane". User ah1-993's metapragmatic insertion "- *you shouldn't* really recommend something and *say* it helps 'arthritis' without being specific-" in reply 2 can be deemed as a critique on user ah1-365's discourse in reply 1. Furthermore, the other *you* in ah1-993's reply "*you* can't really know if it is the vinegar that is helping" may have been interpreted by ah1-365 as targeting them, rather than a generic *you*, as evidenced by their reply 3 as discussed below. Both critiques in ah1-993's reply may have triggered clarification by user ah1-365 who uses meta-language to emphasise their epistemic status and viewpoints. The metapragmatic discussions between the two users thus ensue.

In reply 3, ah1-365 clarifies their epistemic status "I know how", "I also know" and repeats what they have mentioned in reply 1 "Everyone is different...". Although there is no hedge or modal, the latter can be seen as a qualification of their argument. This qualification is highlighted repeatedly by this user with the word *Note* and communicative verbs "I also clearly stated" in reply 3 and "I also *said*" in reply 5 at the end of both replies. In reply 5, probably after ah1-993 continues critiquing in reply 4 with eight "*you*" directed towards ah1-365 (see paragraph below for elaboration), ah1-365 spells out the implicature of "Everyone is different...", "to me, says a person will think 'might be worth looking into to see if it may help me'."

However, ah1-365's clarification and qualification seems to be dismissed by ah1-993 who also clarifies themselves, as evidenced in reply 4, "I *wasn't* querying what *you said*", "I *didn't say* that *you didn't* know.....or that *you don't* know", "*you* made a blanket recommendation for all", "The *point* I was trying to make". However, ah1-993 seems to be furthering critiques towards ah1-365 by referring to another comment made elsewhere, "In another *post*, about apple cider vinegar, *you* specifically *say* 'Two words - Take It.'" and current thread "In this *post* *you* tell us 'trust me, it works'. That sounds to me as *if you* are recommending it to anyone who reads that *comment*." Ah1-993 ends their clarification with a question "Don't *you* think that on an academic website we should

discuss ...". This question starts a metapragmatic discussion on what should be written in the online discussion in FutureLearn. It is worth noticing that ah1-993's increasingly face-threatening replies, i.e., from current thread to another thread and metapragmatic comment, can be a way of continuing the conversation in response to the repetitions of the same comment by ah1-365. Such dynamic was also found by Marra (2012) in face-to-face disagreement and can be a sign of stalemate.

The metapragmatic question in fact is agreed by ah1-365 "*Yes, I agree, academic websites are all about discussions and debates*". However, ah1-365 also raises their metapragmatic concern on writing comments in online discussions, as evidenced by their reply to ah1-993, "*but you don't have to post long, in-depth comments on everything*". In fact, similar comments towards ah1-993's commenting have also been made elsewhere by other users, probably because ah1-993 has contributed 361 replies in this course. The agreement expressed by ah1-365 as well as identity performance, "*my enthusiasm can get the best of me*" can also be a concession strategy, but ah1-365 also voices their view on how others might utilize online information, "*Maybe I should not give people credit to not blindly take something that a stranger has enthused about.*", "*I know very few people who go on a site and don't follow other links about that product and each site you go on you find out more and peoples' opinions albeit good or bad.*"

8.5.4.5 End of the thread

Although user ah1-365 has engaged in clarification and concession, this thread does not seem to end with a reconciliation, as evidenced by ah1-993's continued critique towards ah1-365 with at least four *you* directed to them (Figure 8.12).

Figure 8.12 End of thread 20392679

Reply 6 (subsequent contribution) 2017-03-23 09:49:24 Likes: 0

User ah1-993

Sorry [ah1-365], **but you** seem to be completely missing my **point**.

Apple cider vinegar, taken in reasonable amounts is likely to be harmless. I consume it myself (although not for its health benefits!)

I was not trying to have an academic discussion about whether or not people are sensible (although in fact **you don't** know who is reading **your** posts and there are many reports of people being harmed by taking something that has been recommended over the internet)., **but** about whether or not apple cider vinegar could be **said** to be good for arthritis and whether or not one person's anecdote is good evidence of efficacy.

No, I **don't** think **your reply** was flippant, **but**, as I **said**, **you can't** really know **if** it is the vinegar that is helping or **if** this is **just** a natural fluctuation in the condition. **You** have never **replied** to my original **point**, **but** that would make a good academic discussion! Or perhaps we could discuss this in reference to cayenne pepper **if you** find that less emotive ;-)

In the last reply in this thread, ah1-993 starts their comment with keyword *sorry*, but it does not seem to be a genuine apology. Instead, it is used with another keyword *but* and this collocation pattern is a common disagreement token (Baker, 2014; Baym, 1996). This is evidenced by “*you seem to be completely missing my point*” which follows “*sorry...but*”. This strategy of disagreement seems to be used again towards the end of the reply where user ah1-993 first seems to be hedging, “*No, I don't think your reply was flippant, but...*”, which is followed by critique towards ah1-365, “*You have never replied to my original point, but that would make a good academic discussion*”. Besides this continued metalinguistic comment on other's comments, ah1-993 also continues metapragmatic discussions, “*in fact you don't know who is reading your posts*” despite they themselves mentioning “*I was not trying to have an academic discussion on whether or not people are sensible*”. The repeated emphasis of the discussion to be “*academic*” may constitute a value judgment on others' comments. User ah1-365 never comes back to the thread, despite ah1-993's invitation “*perhaps we could discuss this in reference to cayenne pepper if you find that less emotive;*”) with an emoji.

8.6 Conclusions regarding micro-analysis and the importance of *agree to disagree*

The micro-analysis of three threads in this section focuses on how users negotiate with each other when disagreement arises, thus explicating discourse practices underlying their interaction within a thread. The three threads analysed mainly comprise dialogue between two disagreeing users.

Methodologically, because of their continuous engagement, I am able to interpret each turn based on what they have contributed before and after. Overall, the conclusions highlight the importance of discursive negotiation around disagreement to establish intersubjectivity, even when this does not lead to a convergence of views. This is further illustrated in the final section of this chapter where the communicative functions of *agree to disagree* is examined.

Before reaching this conclusion, however, I will discuss the discourse practices underlying the processes of intersubjectivity between users who take different stances in the three threads. In the first two threads, the disagreeing users negotiate with each other and converge to mutual understanding whereas in the third thread, users remain in stalemate towards the end of the thread. Although all three threads consist of stance-taking between users, similarities and differences in the discourse practices employed for negotiation can be identified. It can be argued that the discourse practices employed in the first two threads facilitate a dialogic conversation, whereas the conversation in the third thread seems to be parallel monologues. These are explored in the following subsections.

8.6.1 Concession with *but*

As shown in the first two threads where users come to an understanding with each other, conceding others' points then reasserts one's own point seems to expand the dialogic space to accommodate multiple voices. The reassertion typically incorporates others' points and establishes qualification with *if*-conditionals. This facilitates intersubjectivity as both parties are taking in others' arguments

while updating their own arguments. This concession is in line with Lindström & Londen's (2013) suggestion that concession is an important strategy for reasoning and argumentation.

However, concession does not necessarily facilitate intersubjectivity when it is just a “show”, as argued by Antaki & Wetherell (1999). This can be observed in the third thread where the reassertion introduced by the concession involves criticism or simply repetition of one point without reformulation after the other’s replies. The criticism is evidenced by *you* that has been repeatedly used in the reassertion, and there is no qualification, hedging or reformulation of stances.

8.6.2 Meta-language for clarification and criticism

Similar to concession, meta-language realizes discourse practices differently. This difference can be observed from the second and third thread. In the second thread, the clarification of what have been said and understood is well-taken by the other party. Both users express mutual understanding and attribute epistemic authority to each other by acknowledging each other are possibly right to some extent under the conditions they specify, thus engaging in intersubjectivity. This is in line with previous research that suggest that meta-communication is a realization of the reflexive capacity of language in human communication (Swales, 2001; Tanskanen, 2007), and facilitative group collaborations (Stahl, 2015) and mediation (Janier & Reed, 2017). Meta-language also creates common grounds when interlocutors point out the similarity or differences in their views (Liu & Liu, 2017), thus delineating differences underlying a conflict and making explicit what is at issue in the discussions.

However, in the third thread, despite one user’s repetition of their hedged argument and use of meta-language to emphasise what they have and have not said, the other user seems to be continuing criticism towards their comments and discourse, and to some extent aggravates it by referring to their other comments elsewhere. Frequent usage of *you* to refer to each other in disagreements, alongside negative evaluations of each other’s discourse or epistemic status, suggests that each user maintains their own subjectivity despite clarification from others, and can

risk being interpreted as hostile (S. Scott, 2002; Sotillo & Wang-Gempp, 2016). Furthermore, the clarifications in the third thread seems to lead to metapragmatic discussions, whereas in the second thread the discussion is still on the content.

8.6.3 Metapragmatic expressions

As reviewed in Chapter 3, metapragmatic expressions are judgements regarding the appropriateness of one's own or others' posting (Kleinke, 2010; Tanskanen, 2007). In the third thread, the metapragmatic expressions are negative evaluations. Similar metapragmatic expressions have also been found elsewhere in the corpus in terms of norms of learning and posting in MOOCs. Given the evaluative nature of one's posting behaviours, they could easily be interpreted by interlocutors as personal attack and challenge (Angouri & Tseliga, 2010; Bou-Franch & Garcés-Conejos Blitvich, 2014; S. L. Graham, 2007; Janier & Reed, 2017; Jenkins & Dragojevic, 2013; Kleinke, 2010; Rees-Miller, 2000; S. Scott, 2002). Although there is no obvious flaming in this thread, this thread does not seem to converge to mutual understanding, compared to the other two threads. This suggest that, although meta-language could be helpful to establish shared understanding for negotiation during disagreement (Liu & Liu, 2017; Nathan et al., 2007; Robinson, 2009), there is a risk for aggravating disagreement if it is used for repeated negative metapragmatic expressions towards others and asserting one's own points without considering the presence of alternative voices by others (Jenkins & Dragojevic, 2013; Kleinke & Bos, 2015).

8.6.4 Putting together keyword analysis and micro-analysis of threads

The keyword analysis and micro-analysis of threads attest to the complexity of stance-taking in the online discussion of FutureLearn. The analysis in particular highlights the importance of discourse practices used not to reconcile, but to acknowledge, entrenched opposing stances. Although agreements are more common, disagreements seem to be where users engage in continued and sustained interactions. In these sustained interactions, users employ various discourse practices in their negotiation to achieve intersubjectivity and explore different voices. This speaks to the

potential importance of disagreement in online discussions, as agreed by various scholars (e.g., Lapadat, 2007; Lewiński, 2013; Littleton & Whitelock, 2005; Marra, 2012; Mercer, 2004). The finding thus far also suggests that users holding different views are brought together in the FutureLearn discussion space, and could engage in negotiation, instead of an echo chamber with similar voices as identified in other online spaces (e.g. Freelon, 2015; Veletsianos, Kimmons, Larsen, Dousay, & Lowenthal, 2018; Walter, et al., 2018).

Importantly, disagreement does not necessarily lead to convergence, and users can still hold on to their stances after being challenged by other users. Despite not agreeing with each other, disagreeing parties are at least exposed to alternative voices and engage in continual negotiation to reach mutual understanding of each other's view, and possibly reformulation of one's own stance (Marra, 2012; Mercer, 2004; Nathan et al., 2007; Sarewitz, 2011). It can be argued that this continual negotiation is a dialogic conversation because different voices are in a dialogic relationship. However, disagreement can become unproductive argument when disagreeing parties blame each other, pursue to "win" over the other, or ignore others' emotions or words (Berryman-Fink, 1998; Felton et al., 2015; S. L. Graham, 2007). The third thread in the micro-analysis illustrates this situation where users continuously critique other's posting behaviours despite the other's clarifications attempted with meta-language. Admittedly, it is not possible to determine whether users are simply trying to 'win' over each other through this analysis. However, based on the users' textual contributions, it can be argued that the conversation in the third thread can at times be parallel monologues, as users continue their argument without acknowledging others' voice or meta-comments. This raises the important question as to how users who cannot reach an agreement after negotiation might find a way to recognize their entrenched differences, such that everybody's voices are still acknowledged, and the dialogic space is still open. This is explored below in the final analysis of this chapter.

8.7 Agree to disagree

To explore users' social behaviours and discourse practices when they cannot agree with each other after discussions, a micro-analysis of threads containing the phrases *agree to disagree* and *agree to differ* is conducted. Investigation of the communicative functions of these phrases may also reflect discourse practices in other similar situations that do not use these exact phrases. This choice is informed by my preliminary reading of the long threads in the corpus and literature review on disagreement. The search for *agree to disagree* results in 10 counts across nine threads, and *agree to differ* in four counts in four threads, all of which are in subsequent contributions in the threads, suggesting that users use this phrase after some exchange with others. There are another two instances in initiating post and independent post, but they are used in a quote, rather than being used to address other users, so they are not considered in the current analysis.

Although the frequency of *agree to disagree/differ* is low, it is relatively more frequent when compared to other general corpora, as shown in Table 8.6. The observation that these phrases are used more frequently in online discussions, in both my corpus and the Yahoo News Annotated Comments Corpus (Napoles et al., 2017) which consists of threads of argumentative nature in online news websites, suggests the importance of these phrases in online discussions. The investigation of the social behaviours and discourse practices where these phrases occur may thus reveal user-user interactions characterizing online discussions. Besides the dictionary meaning of “*stop arguing*” (Collins Dictionary, n.d.), i.e., intended to function to bring the discussion to a close, it is found that FutureLearn users also use these phrases for a range of negotiating functions. Before examining these functions, the interaction patterns of the threads where these phrases occur are first introduced to give an overview of the context.

Table 8.6 Normalized frequency in per million words of agree to disagree and agree to differ in different corpus

Corpus	<i>agree to disagree</i>	<i>agree to differ</i>
FutureLearn online discussions	1.07	0.35
Yahoo News Annotated Comments (Napoles et al., 2017)	1.85	0
Spoken BNC2014 (Love et al., 2017)	0.26	0
Written BNC1994 (Leech et al., 2001)	0.05	0.05

8.7.1 Interaction patterns in the threads

The interaction patterns of the 13 threads containing the mention of *agree to disagree/differ* are examined in terms of the number of replies, involvement of super-posters, and the extent to which the phrases serve to bring the discussion to a close (Table 8.7).

Table 8.7 Interaction patterns in the threads where agree to disagree/differ is used.

course	Number of replies	Super-posters' involvement in the disagreement	Last reply in the thread ¹	Speaker ² comes back	Addressee ³ comes back	Speaker who is the initiator	Thread Id	Figure where threads are shown	Description in section
finance-1	23				√	√	4352189	8.13	8.7.2
finance-1	6					√	4447791		8.7.2
ancient-1	8	√				√	19841865		8.7.2
finance-1	17	√			√		4505886	8.14	8.7.3
finance-1	38	√		√	√		4436943		8.7.3
ancient-1	15		√			√	20327421		8.7.3 + 8.7.4
finance-1	6	√	√			√	4300841		8.7.4
management-4	6		√			√	18030898	8.15	8.7.4
ancient-1	6		√				19422436	8.16	8.7.4
nutrition-4	7	√			√	√	19101192		8.7.4
nutrition-4	24			√	√		19180872	8.17	8.7.5
ancient-1	9	√		√	√		19412152	8.18	8.7.6
ancient-1	26	√		√	√	√	20311486		8.7.6

¹ The phrase *agree to disagree/differ* is used in the last reply in the thread, i.e., nobody replies afterwards.

² *Speaker* refers to the user who uses the phrase *agree to disagree/differ*.

³ *Addressee* refers to the user who is the target of the phrase *agree to disagree/differ*.

Five observations are worth noting. First, all the threads are at least five replies long, and four even consist of more than 20 replies, suggesting that *agree to disagree/differ* and similar situations might occur in sustained interactions. Second, most of the threads are from finance-1 and ancient-1, perhaps because both courses are focused on contentious issues. Third, one super-poster from each of these two MOOCs are involved in three threads each. They are not the one saying these phrases but being addressed to. Based on my readings of the comments in the corpus, and quantitative analysis of their replying patterns, these super-posters seem to be very responsive to others' replies and vocal with their opinions. They also contribute proportionally more replies than new posts¹⁴². Fourth, in eight out of the 14 threads found, the initiator is the one who says this phrase, suggesting that they have not changed their original view, although they do come back to the thread, compared to those initiators who just post and go. The fact that the initiator comes back suggests engagement which allows for the possibility that they have at least taken on board the alternative viewpoint. Fifth, the phrase does "stop" the argument. In some occasions, it indeed occurs in the last reply. In other occasions, the speaker or the addressee does not come back to the thread anymore, although others continue the thread. However, there are also occasions where conversation continues between the speaker and the addressee of the phrase. This heterogeneous interaction pattern suggests complicated social behaviours and discourse practices in these threads where disagreement arises. They are explicated in the following subsections by presenting selected threads, while other similar threads are indicated in Table 8.7 above. In the analysis below, the focus is on *agree to disagree/differ* only, so only the relevant part of the thread is shown. This contrasts with the micro-analysis in section 8.5 which analysed how multiple discourse practices are used throughout a thread.

¹⁴² The super-poster in the ancient-1 course contributed 361 replies and only 8 posts, whereas the one in the finance-1 course contributed 302 replies and 81 posts.

8.7.2 Acknowledging disagreement without negotiation

In three threads, each initiator does not engage in further negotiation but only reply with *agree to disagree* after the other users express disagreement towards their initiating post, as illustrated in Figure 8.13.

Figure 8.13 Part of thread 4352189.

Thread 4352189

Source: <https://www.futurelearn.com/courses/inequalities-in-personal-finance/1/comments/4352189>

Initiating post 2015-03-30 09:25:46 Like: 0

User f1-303

I'm afraid that there too many people coming into the UK and probably for low skill low pay jobs and either not paying or not paying enough into the system and generally being a drain on all resources [.....]

Reply 1 (first contribution) 2015-03-30 10:22:44 Like: 21

User f1-359

[User f1-303], I have to disagree with your comments on people going to the UK (to work, low paid or otherwise) being a drain. They pay into the system via taxation and NI contributions. I know a number of Polish, Hungarian, Bosnian, Serbs and Croatians who have worked in the UK [.....]

Perhaps you are thinking of the illegal elements, within the UK society, that exploit those less fortunate than others? Those illegal elements usually use slave labour and the unfortunate individuals are hidden from society with no access to state support.

Reply 2 (initiator's subsequent contribution) 2015-03-30 11:06:01 Like: 0

User f1-303

[User f1-359], we will have to agree to disagree.

Reply 3 (first contribution) 2015-03-30 14:43:11 Like: 3

User f1-544

I *agree* with you [User f1-303].

The current personal allowance is £10,600 and with tax credits, there's no net tax income to the treasury until an income of over approx. £18,000 is achieved - and by definition, £18,000 isn't low paying (it's not exactly high paying, but it meets the "living wage"). [.....]

[Reply 4 to Reply 11 are omitted]

Reply 12 (subsequent contribution) 2015-03-31 00:25:23 Like: 3

User f1-359

[User f1-303], *No* problem, it's a debating *point* and we both have our views.

[User f1-544], I have just checked the HMRC web site and your figures are incorrect, although not a million miles away [.....]

[There are 10 more replies afterwards]

Note. Agree to disagree/differ is underlined for emphasis.

In the thread in Figure 8.13 and another two threads not shown here, the initiators post their stance which receives explicit objection from others, for example “I have to disagree with your comments” in Figure 8.13 above. After the disagreeing replies, the initiators of two threads reply with *agree to disagree* only without any additional content to explain their stance, for example “we will have to agree to disagree” in Figure 8.13. In the third thread (thread 19841865), the initiator replies once to the disagreement with an explanation but receives other objections afterwards, then resorts to just an expression of *agree to disagree* in their last contribution in the thread. In all these threads, after the *agree to disagree*, the threads continue with at least one reply from other users. In Figure 8.13, the user being asked to *agree to disagree* also acknowledges the initiator in reply 12 by saying “No problem, it’s a debating *point* and we both have our views” but continues discussions with others.

It seems that these initiators read the others’ views but do not engage in negotiation. This is evidenced by their use of *agree to disagree* to acknowledge others’ views and withdrawal from the thread. This is despite the fact that there are also other users supporting their point of view, for example reply 3 in Figure 8.13, “I agree with you”. Given that they do not provide elaboration in their reply, it might be hard for the other users to respond to them. Nonetheless, their withdrawal does not prevent other users from continuing the discussions, given the polylogal nature of threads.

In other threads to be examined in the following subsections, users raise *agree to disagree/differ* after at least some exchange with others. Most of these replies also contain elaboration besides *agree to disagree/differ*, unlike the discourse practice of not engaging in any negotiation and simply voicing *agree to disagree/differ*. Based on users’ discourse surrounding the *agree to disagree/differ* in their replies, three other communicative functions can be drawn: reconciling, summarizing differences following negotiation, framing on-going discussions for further exchange.

8.7.3 Reconciliation

In three threads, disagreeing users explicitly *agree to disagree* after some exchange, suggesting they leave on good terms, as illustrated in Figure 8.14.

Figure 8.14 Part of thread 4505886.

Thread 4505886
Source: <https://www.futurelearn.com/courses/inequalities-in-personal-finance/1/comments/4505886>

[initiating post and reply 1 to reply 14 are omitted]

Reply 14 (subsequent reply) 2015-04-24 16:32:38 Likes: 1
User f1-456
[f1-247] *you* are mussing the *point*..my parents like [f1-348] saw the house ad tge mist important thing so we as children missed out such as holidays and nice clothes compared with friends [.....]

Reply 15 (subsequent reply) 2015-04-24 18:11:23 Likes: 0
User f1-247
Hi [f1-456] - I think *you're* missing the *point*. *But* I'm finishing this course now, and moving on to the next. We'll have to agree to differ. Regards, [f1-247].

Reply 16 (subsequent reply) 2015-04-24 18:22:48 Likes: 1
User f1-456
No way am I missing any *point* *but* I really do wish *you* well with *your* next course..hopefully less contentious! ..I am having a month off before my next one

[one more reply from another user is omitted]

The disagreement between users f1-456 and f1-247 starts from reply 7 (not shown here) and two of them have been addressing each other before they reach the conversation shown in Figure 8.14. Although both users still critique each other for “mussing the point” and defend themselves, “*No* way am I missing any *point*”, they express good will to end their last reply to each other, as evidenced by “Regards”, “I really do wish *you* well”. The declaration of f1-247 that they are “finishing this course” and “moving on” may also serve as leave-taking to indicate they are not coming back to the thread. The user f1-456 seems to reciprocate this way of ending their discussion by also

mentioning what they are going to do next, “I am having a month off”. The disagreement thus shifts to a friendly ending after some negotiation.

8.7.4 Summarizing differences following negotiation

In six threads, *agree to disagree* is raised by a user in their last contribution in the thread after engaging in some exchange with other users. In these replies, following the use of *agree to disagree/differ*, the users continue elaborating on their stances, as shown in Figure 8.15 and Figure 8.16.

Figure 8.15 Part of thread 18030898.

Thread 18030898
Source: <https://www.futurelearn.com/courses/contract-management/4/comments/18030898>

Initiating Post 2016-11-22 Likes: 6
User m4-67
I am always very sceptic about the "vision and mission"-terms. [.....]

[Reply 1 to reply 2 are omitted]

Reply 3 (first contribution) 2016-11-23 Likes: 0
User m4-435
[m4-67], are *you* suggesting that there *isn't* a need for a vision or a mission because ultimately the motives of a business/entrepreneur are financial? I'm not sure that I understand *your* point and would like some clarity. I think having a vision or mission is critical, otherwise, what are *you* working toward? [.....]

Reply 4 (initiator's subsequent contribution) 2016-11-23 Likes: 0
User m4-67
How will having a vision influence what *you* are doing, please? Have *you* never asked yourself after reading a "Vision Statement" "Now, where have I read that before? Was it not on the website of the competitor???" I always feel that such statements ring hollow [.....] YES I do believe that there is no need for a vision.

Reply 5 (first contribution) 2016-12-08 Likes: 0
User m4-844
I *agree* with *you* to some extent [m4-67], *but* in my opinion those organisations who don't have a vision or mission statement generally will not have a clear goal to work towards [.....]

Reply 6 (initiator's subsequent contribution) 2016-12-08 Likes: 0
User m4-67
I think I will have to agree to disagree with the people who believe vision statements are useful. *Maybe* that is because I am a lawyer by profession (i.e. I am being paid to be sceptical) or *maybe* it is because I am just a *sorry*, jaded cynic. Whatever it is, to me this sounds only like the invention of some clever business consultant who dubbed a phrase that everybody uses because they are afraid to cry "the emperor is naked".

In Figure 8.15, the initiating post attracts agreement in reply 1 and possibly reply 2 (not shown here) but disagreement in reply 3, which prompts the initiator to defend their position and emphasise their stance in reply 4, before another objection is raised in reply 5. In the last reply of the thread, reply 6, the initiator suggests they “will have to agree to disagree”, before acknowledging

the existence of different views held by “people who believe vision statements are useful”. At the same time, multifaceted identity work, “*Maybe* that is because I am a lawyer by profession (i.e. I am being paid to be sceptical) or *maybe* it is because I am just a sorry, jaded cynic”, seems to be used as possible explanations underlying the differences in stances, signalled by the repetition of “*maybe ...because*”. Therefore, the entrenched disagreement is attributed to identity differences, which are not negotiable, rather than to debateable arguments or evidence. In light of this identity work, the initiator m4-67 further provides an argument which is framed based on identity, “some clever business consultant who dubbed a phrase”. It is also worth noticing that in this thread, only the initiator makes subsequent contributions to the thread, while other replying users do not continue engagement. It can be considered as a one-to-many polylogue.

Similarly, in Figure 8.16, the user who mentions *agree to disagree* also attributes their disagreement to a non-content issue.

Figure 8.16 Part of thread 19422436.

Thread 19422436
Source: <https://www.futurelearn.com/courses/ancient-health/1/comments/19422436>

[initiating post, reply 1 to reply 2 are omitted]

Reply 3 (first contribution) 2017-02-07 13:20:38 Likes: 2
User ah1-198
People seem to think we have a 'right' to good health and therefore someone else has a duty to provide it for us. I wonder if [ah1-240] would be so generous *if* it were us stealing from someone else's national health service?

Reply 4 (subsequent contribution) 2017-02-09 05:03:03 Likes: 0
User ah1-240
[ah1-198] - yes I believe that in a civilised society like ours citizens have a right to good health, amongst other things, and also an entitlement based on their responsibilities which include paying fair taxation and helping others less fortunate in society than ourselves [.....] Abuse of any system is not to be encouraged *but* let us employ some intelligence and a relative perspective and not witch-hunt any particular group. Dunno about 'anyone else's health service' *but* just in terms of British imperial and colonial history there are plenty of examples of 'stealing' by the British Empire. [.....]

Reply 5 (subsequent contribution) 2017-02-09 18:55:43 Likes: 1
User ah1-198
[ah1-240] it's nonsense to talk about a 'right' to good health. Who are *you* going to sue when *you're* ill?

I don't see how *you* arrive at condemnation of thieves being a witch hunt.

Nor does it make sense to say that people today owe a debt for actions of people in the distant past, even *if your* wild generalisation about the imperial project were valid. *You* do know there was a net outflow of capital from the UK to the colonies during the period of empire don't you? It cost money then and there's *no* reason it should continue to cost money now.

Reply 6 (subsequent contribution) 2017-02-10 10:41:55 Likes: 0
User ah1-240
[ah1-198], I suspect our world views and politics are polar opposites and we will never *agree*. We even seem to talk different languages. Health is a relative concept - not an absolute one. A society which promotes health for its citizens to an optimum, with those citizens taking individual responsibility in the personal sphere as much as possible seems desirable to me. Where suing comes into it I can't imagine? Let us agree to disagree and move on. Bye, bye.

The thread in Figure 8.16 is started by a user who never comes back to the thread. Reply 1 is also contributed by a user who never comes back, although it receives reply 2 from user ah1-240 who disagrees with them. After ah1-240's reply 2, the thread develops into a one-to-one dialogue

between user ah1-240 and ah1-198 who disagree with each other, as evidenced by their addressing each other's name in their replies. In the last reply of this thread, user ah1-240 seems to attribute the disagreement to their entrenched difference in "world views and politics" and highlights they "are polar opposites" and "will never *agree*". Similar to Figure 8.15, as well as other threads not shown here, user ah1-240 also elaborates their stance with more content, "health is ...", in their *agree to disagree/differ* reply.

It is possible that *agree to disagree* is used to frame their reply as the last word on the subject, by acknowledging others' views while summing up their differences and reiterating their own argument. This wrapping up function is also evidenced by phrases such as "move on", "bye, bye", as shown in Figure 8.16, and "There's more important stuff to talk about, anyway", "One last word from me", "Enough said – let's move" in other threads not shown here. It appears that in most cases the other user obliges and does not come back to the thread, as *agree to disagree* appears in the last reply in five out of the six threads. In short, in this scenario, users typically summarize their own stance in their *agree to disagree* reply, while some users also attribute the disagreement to identity or ideological differences, which to some extent might be seen as irresolvable.

8.7.5 Framing on-going discussions

In contrast to threads presented in preceding subsections, disagreeing users in three threads still continue their discussion with each other after *agree to disagree/differ* is raised. In one thread, the phrase *agree to differ* seems to be used to frame their on-going discussion as a way to explore ideas, rather than to stop their discussion, as shown in Figure 8.17, while the other two threads as will be discussed in the next subsection shows on-going disagreement.

Figure 8.17 Part of thread 19180872.

Thread 19180872
Source: <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/19180872>

[Initiating post and reply 1 to reply 13 are omitted]

Reply 14 (subsequent contribution) 2017-01-29 21:32:09 Likes: 1
User n4-3057
[n4-1188] - I think we will need to *agree to differ* on many ideas. lol.
I *agree* though that fats are not the demons they were once thought to be. It is sugar that is the issue - with high fructose sugar being the culprit. See this interesting lecture
<https://youtu.be/dBnniua6-oM>
The animals we slaughter for meat can also have these hormones in their muscle tissue. [.....]

Reply 15 (initiator's subsequent contribution) 2017-01-29 22:19:59 Likes: 2
User n4-1188
Thanks for the the *link*. I *agree* about sugar; *but* more specifically about the high fructose corn syrup and the other liquid sugar abominations. [.....]

Reply 16 (subsequent contribution) 2017-01-29 23:25:58 Likes: 1
User n4-3057
I've just finished watching the video 'sugar. the bitter truth' that I posted in an earlier message and am stunned. [.....]

Reply 17 (initiator's subsequent contribution) 2017-01-30 00:22:36 Likes: 2
User n4-1188
Well, I guess there are just too much money to be made from selling soft drinks [.....]

In Figure 8.17, *agree to differ* appears in reply 14, which is in the middle of the exchange between users n4-3057 and n4-1188, where they disagree once, and each contributes at least three times before this reply. They continue for another three direct turns after the *agree to differ* reply. User n4-1188 starts reply 14 with *agree to differ* coupled with “lol” before expressing their view. This reply does not seem to be used for summary or reconciliation, but is used to frame the discussion as exploratory talk, as evidenced by the expression of agreement, “I *agree* though” and an invitation to see a video, “See this interesting lecture”. In turn, user n4-1188 acknowledges the video, “*Thanks* for

the *link*” and expresses agreement, “I *agree* about sugar”, towards user n4-3057 who had previously stated “It is sugar that is the issue”.

8.7.6 Failing to *agree to disagree*

The *agree to disagree/differ* conversations examined thus far have illustrated that users recognize their differences, either disengage or wrap up their disagreement, or continue exploring each other’s view. However, there are two exceptional threads in which a user raises *agree to disagree* appeal two times before both the user and the other disagreeing user wrap up their disagreement. It happens that these two threads involve one same super-poster who continues voicing their disagreement after the first *agree to disagree* is raised. It is possible that disagreeing users in these threads argue to “win” a debate such that they do not recognize that a difference can exist and it does not need to be resolved, unlike those shown in the previous threads. This is despite both disagreeing users acknowledging to *agree to disagree*, as illustrated in Figure 8.18.

Figure 8.18 Part of thread 19412152.

Thread 19412152

Source: <https://www.futurelearn.com/courses/ancient-health/1/comments/19412152>

[initiating post and reply 1 to reply 20 are omitted]

Reply 21 (subsequent contribution) 2017-02-08 17:18:04 Likes: 0

ah1-622

Think we will agree to disagree

Reply 22 (subsequent contribution) 2017-02-08 18:08:06 Likes: 5

ah1-993

Sadly, we probably will.

I'm not religious, *but* I do believe in 'Love thy neighbour'.

That's easy to do when our neighbours are very similar to us. A lot harder when they are seen as different. I believe in support, not blame.

Reply 23 (subsequent contribution) 2017-01-30 00:22:36 Likes: 2

ah1-292

I *agree* with you [ah1-993]. I'm *no* Doctor *but* I do think addiction is a form of mental illness and affects some surprising people!

Reply 24 (subsequent contribution) 2017-02-09 06:49:23 Likes: 1

ah1-622

over eating, smoking, taking drugs are life style choices that then become addictions.

'Love the sinner, but do not condone or tolerate the sin'...we are all sinners

Reply 25 (subsequent contribution) 2017-02-09 08:24:09 Likes: 2

ah1-993

A great quote [ah1-622], *but* what a shame if mental ill health is seen as a sin! Doesn't that take us straight back to the middle ages? And how come we are urged to love the sinner...just not enough to give them medical care?

The problem is [.....]

Reply 26 (subsequent contribution) 2017-02-09 08:41:10 Likes: 0

ah1-622

This disagreement can run and run *if you* want, I don't, as this is neither place nor time. Mental health issue can be numerous, genetic or due to lifestyle choices...nothing is 'black or white'.

It is a grave mistake to use mental health issues/poor child hood etc etc as excuses for unhealthy lifestyle choices and then expect society and NHS to put things right without any input from patient.

Oh, and to clarify [ah1-993]...I am not 'religious', just some one who has a born again faith,

Reply 27 (subsequent contribution) 2017-02-09 10:43:34 Likes: 1

ah1-691

I must *agree* with poor [ah1-622] who has to write on and on about this problem while she does not want to. But still I want to add my bit. When somebody smokes or drinks too much all their life, I am of the opinion that *you* only have to thank yourself for the illnesses *you* get in later life [.....]

Reply 28 (subsequent contribution) 2017-02-09 11:13:23 Likes: 2

ah-993

[ah1-622], *you* seem to be misunderstanding what I *say*. Sometimes childhood trauma is a reason why people have mental health issues later in life (something backed up by research), *but* I didn't *say* that is an excuse to live however *you* want. And I certainly never *said* that people should be put right without input from the patient (that input would be essential), all I *said* was that people needed medical support, sometimes that is vital. After all, alcoholics can, and do, die if they suddenly stop drinking.

And I didn't *say* anything about anyone's religion, I simply *said* (as I was using a biblical quote) that I was not religious.

Oh dear [ah1-691] - I am not making anyone write on and on about a topic *if* they don't want to - no-one has to join in or continue, *if* they don't want. But on FutureLearn we have discussion forums so that people can discuss the topic.

I know that *your* viewpoint is common and I am probably in a minority, *but* I am saddened when people think that mental ill health is not an illness and is not worthy of medical support.

[reply 29 to reply 33 are omitted]

In Figure 8.18, users ah1-622 and ah1-993 have been engaging in stance-taking in replies 12, 13, 18, 19, 20 (not shown here) before user ah1-622 raises *agree to disagree* in reply 21, which may be a leave-taking, similar to other threads shown previously. Although user ah1-993 reciprocates by saying “Sadly, we probably will” in reply 22, the exchange between the two users continues in reply 24, 25, 26, 28. There are a number of potential explanations for this continuation. Firstly, it is perhaps because another user ah1-292 aligns with ah1-993 in reply 23, “I *agree* with you [ah1-993] ...”, thus dis-aligning with ah1-622. This might have attracted ah1-622 to come back to defend their position in reply 24. A similar observation is also made in another thread not shown here where the (dis)alignment of other users with either of the disagreeing users triggers them to come back to the thread. Secondly, it is also possible that ah1-993 still puts forward new ideas in response to ah1-622, “I’m not religious, *but* I do believe in ‘Love thy neighbour’....”, that might attract ah1-622 to come back, as evidenced in the parallelism in their reply “Love the sinner”. This is also observed in another thread not shown here. Thirdly, the exchange continues probably because both are prolific contributors in the course, ah1-622 contributes 156 comments whereas ah1-993 contributes 368 comments. Lastly, before *agree to disagree* is raised, these two disagreeing users have been contributing to the thread for seven times each. It might be harder to terminate an on-going extended discussion.

Whatever the reason, their exchange continues and seems to shift to metapragmatic discussion about how the FutureLearn discussion should be. This shift happens when user ah1-622 reiterates in reply 26 “This disagreement can run and run *if you* want, I *don’t*, as this is neither place nor time”. This reply can also be interpreted as the second time the user ah1-622 appeals to *agree to disagree*, as evidenced in another user ah1-691’s reply 27. In this reply the user aligns with ah1-622, “I must *agree* with poor [ah1-622] who has to write on and on about this problem while she does not want to. *But* still I want to add my bit....”. This support seems to attract ah1-993’s reply 28 “...Oh dear [ah1-691], I am not making anyone write on and on about a topic *if they don’t* want to -

no-one has to join in or continue, if they don't want. But on FutureLearn we have discussion forums so that people can discuss the topic...".

This metapragmatic discussion in Figure 8.18 shows that some users, for example ah1-622, consider long-running disagreement should stop at some point, and signal this understanding with *agree to disagree*. At the same time, other users, for example ah1-993 adopts a slightly different view that people can assert their agency to join or leave a discussion, leaving others to voice their opinions. This perspective is also evidenced in another thread not shown here where the same user ah1-993 asks other users in the thread “*don't read the replies if it annoys you ;-)*” when there is a repeated disagreement between them and another user who has already suggested they *agree to disagree/differ*. This metapragmatic discussion demonstrates that users construe the communicative norms of FutureLearn differently, and thus differ in their perceptions as to whether and when a disagreement should be stopped.

8.7.7 Conclusions regarding *agree to disagree*

The phrase *agree to disagree/differ* maybe a particular characteristic of online discussion spaces, as indicated by the fact that their frequency is relatively higher in the current corpus and Yahoo news comment corpus (Napoles et al., 2017), compared to BNC written and spoken English corpus (Leech et al., 2001; Love et al., 2017). The threads where *agree to disagree/differ* is raised, although only 13 in the present corpus, may point to other threads in which users are similarly polarised but do not use the particular phrase. Based on the analysis of the communicative functions of this phrase, it appears that there are three practices that users who disagree may employ: (1) disengage without negotiation, (2) negotiate then reconcile while maintaining differences, (3) continue disagreement. These three possibilities are discussed in the following subsections. Importantly, the mention of *agree to disagree* suggests that users recognize the existence of alternative views, although only some of them are willing to explore others' ideas, which in turn point to the role of discourse practices in processes of intersubjectivity when users have opposing stances.

8.7.7.1 Disengage from discussion

Some users post a comment, and then come back to the thread to leave a reply with only *agree to disagree/differ* after receiving replies disagreeing with them. According to Bou-Franch & Garcés-Conejos Blitvich (2014), this withdrawal does not resolve disagreement. The withdrawal function of *agree to disagree* is consistent with previous findings in online asynchronous settings where users are free to withdraw from discussions (Herring, 1999). They could even leave without saying a word. This probably explains the quantitative analysis in Chapter 5 that shows relatively few subsequent contributions in the corpus. The freedom to withdraw is consistent with the prompt-focused behaviour where users just post and go.

Nonetheless, at least the users come back to the thread, despite maintaining their own views. Although they do not engage the other users in processes of intersubjectivity, they establish interactivity with others rather than ignoring without replying. Importantly, the fact that the users acknowledges other users' stances with *agree to disagree/differ* suggests that they are exposed to alternative views. Fortunately, other users can continue the discussions, given the polylogal nature of online discussions.

8.7.7.2 Reconcile and maintain differences

Agree to disagree/differ appears to enable users to conclude a discussion, even when they are unable to reach a conclusion after some discussions. In these *agree to disagree/differ* replies, users also maintain their own stances. They sometimes attribute their irresolvable disagreement to other underlying differences such as identity and world views. Sometimes the other party reciprocates this conclusion explicitly, sometimes they seem to oblige the *agree to disagree/differ* appeal by not replying further.

In these threads, *agree to disagree/differ* serves to both wrap up their negotiations and express leave-taking from the discussion, possibly also to avoid getting into heated or uncivil discussions. Compared to withdrawing without engaging in any negotiation, these instances indicate that users have tried to engage intersubjectively with each other before accepting that they would

not be able to reach an agreement. The observation that they stop pursuing after some negotiation also suggests that they are participating in the discussion to exchange views rather than to win over others. If they really wanted to change others' view, they would have carried on arguing. Several researchers have found that arguing to reach a conclusion facilitates co-constructions and integration of others' ideas, in this case mutual understanding of each other's subjectivity, whereas arguing to convince or defend may lead to disputational talk (Chiu, 2008; Felton et al., 2015; S. L. Graham, 2007; Mercer, 2004).

8.7.7.3 Individuality and communicative norms in online discussions

However, there are also rare occasions where users ignore the appeal of *agree to disagree/differ* and continue posting their disagreement and stance. This may be due to the individuality of users, which might be hard to verify. Some users might join the discussions to win and convince others, as evidenced by their repeated attempts to critique others or defend themselves without considering others' views or accepting differences. This attempt differs from other users who use *agree to disagree* to acknowledge and explore differences.

The analysis of threads containing *agree to disagree/differ* also suggests that users construe the phrases and the communicative norms of online discussions differently (Tanskanen, 2007). As evidenced by the metapragmatic discussions among users in the threads where users fail to *agree to disagree/differ*, users seem to have different perceptions as to how the online discussion space should be, specifically when a disagreement arise (Danet, 2013; Herring, 2001; Lambiase, 2010; Zhang et al., 2018). Admittedly, there might not be any hard and fast rule to stop their interactions in a thread when disagreement arises, especially when several researchers have argued that sustained interaction is needed for a constructive negotiation process (Gillani & Eynon, 2014; Lapadat, 2007; Tubman et al., 2016). Furthermore, because communicative norms can be fluid and are co-constructed by participating users, ways of managing disagreement varies across communities or even threads where different users are involved (Marra, 2012; Netz, 2014;

Tanskanen, 2007). Nonetheless, more importantly, as shown in the analysis of this chapter, users can be encouraged to employ discourse practices, such as concession, qualification of their arguments, meta-language to clarify similarities and differences to facilitate their negotiation of disagreement.

8.8 Discussion

This chapter explored online conversations by examining users' replies, i.e., the comments that are directed either towards the initiating post or towards replies that are posted before them within the same thread. The users' replies to each other are examined from three angles: general patterns, unfolding threads where users engage in stance-taking, and the discourse practices when users remain polarized. This is achieved by a keyword analysis coupled with micro-analysis of threads and analysis of the phrase *agree to disagree/differ*. In this discussion, I argue that, although sustained interactions are rare, dialogic conversations are possible in online spaces, with users employing discourse practices that entertain others' voices and facilitate intersubjectivity.

When replying to others, users generally engage in interactive discourse, addressing each other, referring to others' initiating post or reply, aligning or disaligning their stance. When taking a stance, expression of agreement is found to be more common than expression of disagreement, and is used more frequently in short threads and in users' first contributions in a thread. Although agreement creates an interactive and supportive online space, this finding suggest that agreement might not be conducive for exploratory talk and intersubjectivity, which require continual negotiations among users (Kellogg et al., 2014; Lapadat, 2007; Littleton & Whitelock, 2005; Paulus, 2006; Rourke & Kanuka, 2007). This finding lead to the potential importance of disagreement in online discussion, despite its negative connotation.

In contrast to agreement, user-user interactions during disagreement seems to be more sustained, instead of short-lived, especially when users engage in negotiation of each other's stance, rather than just withdraw from disagreement. Although less common than agreement, the existence

of disagreement in the online discussion of FutureLearn suggests that different voices are potentially recognized and in dialogic relationship in MOOCs, thus avoiding echo chamber or disputational talk found in other online spaces (Freelon, 2015; Veletsianos et al., 2018; Walter et al., 2018). Being challenged and exploring alternative views has been one goal of online deliberation, as well as socio-constructive learning in education (Dahlberg, 2001; Dennen & Wieland, 2007; Hall, 2010; Laurillard, 2012; Stahl, 2015).

Disagreement provides users with the chance to negotiate with each other to achieve intersubjectivity, i.e., understanding and integrating each other's subjectivity, and co-construction of new understanding, or exposure to alternative viewpoints even though they could not reach convergent view (Lapadat, 2002; Marra, 2012; Mercer, 2004; Nathan et al., 2007; Sarewitz, 2011). These negotiation processes are shaped by users' discourse in their stance-taking with others, as revealed by the keyword analysis and micro-analysis. Strategies found in the current analysis include concession, qualification, meta-language, identity work. On one hand, concession has been used to both acknowledge others' views and voice one's own view, possibly creating coherence and integrating similarities and pointing out differences in their views; *if*-conditionals qualify one's argument to allow alternative scenarios, which could be raised by other users. These discourse practices indicate that users take into account others' point of view while considering their own view, suggesting intersubjectivity (Du Bois, 2007; Stahl, 2015). On the other hand, meta-language has been used to clarify one's stances, epistemic status or understanding of others' views after misunderstanding arises, suggesting an attempt to negotiate each other's subjectivity. Meta-language helps make explicit any underlying differences or similarities between users (Janier & Reed, 2017; Liu & Liu, 2017), and it might be particularly important in online discussions given that other non-verbal cues are missing.

Lastly, users sometimes acknowledge each other's identity, including profession or epistemic status, or attribute their unresolvable differences to their differences in identity. Probably because most users do not know each other in person, identity performances become facilitative of

explaining one's stance and achieving intersubjectivity (Bou-Franch & Garcés-Conejos Blitvich, 2014; Grabill & Pigg, 2012; Jaworska, 2018). All these discourse practices indicate that users recognize the possible differences among them, and may facilitate their (re-)calibration of own's stance and others' stance along the stance continuum. Although they might not reach a convergent point, as in the case of *agree to disagree*, they at least explore alternative views, reconcile or wrap up their disagreement.

However, some discourse practices can be a double-edged sword, for example meta-language. There are times that users in disagreement do not engage in intersubjectivity but drift to critiquing each other's discourse or to metapragmatic discussions, which also happens in other online discussions (Bou-Franch & Garcés-Conejos Blitvich, 2014; Guiller & Durndell, 2006; Kleinke, 2010). The metalinguistic comments on others' comments or identity and epistemic status, especially with negative evaluations and repetitions, can be a barrier to intersubjectivity. These comments might not be favourable for those users being critiqued or those having different voices. Some of them may be deterred from participating in online discussions because of potential conflicts (Littleton & Whitelock, 2005; Marra, 2012).

The findings thus attest to the value of disagreement for online discussions, how discourse practices facilitate intersubjectivity and exploration of multiple voices when disagreement arises, and how some discourse practices and individual behaviours may render disagreement disruptive. In other words, users' discourse practices can make a difference between a user-user interaction that is just a parallel monologue or a dialogic conversation that entertains alternative voices and involves processes of intersubjectivity.

8.9 Conclusion

The investigation in this chapter moved from the potential start to the development of dialogic conversations. The analysis illustrated that in dialogic conversations, users' discourse practices

enable multiple voices to be in a dialogic relationship for intersubjectivity to occur. This shows that dialogic nature of online discourse is not just about replying but specific discourse practices. This also addresses RQ2 regarding how users' discourse practices may sustain or hinder dialogic conversations.

The data-driven approach taken in this thesis revealed the discourse practices in replies by which users agree or disagree with each other, including meta-language, concession, interactive language, thus addressing RQ1 regarding the linguistic features and discourse practices in replies. Most importantly, the findings speak to the empirical importance of stance-taking in online spaces. Informed by the general patterns of replies revealed by keyword analysis, the micro-analysis of threads extended previous studies (Baym, 1996; Kleinke, 2010), which only examine users' expression of (dis)agreement in the replies or adjacent pairs, to the development of a whole thread. The corpus analysis of reply keywords coupled with micro-analysis further contributes to our understanding of some potentially useful discourse practices for developing dialogic conversations even when users disagree. The analysis also demonstrates the feasibility of combining corpus linguistics with micro-analysis to explore user-user interactions in online spaces.

The data-driven approach further revealed one discourse practice that, as far as I am aware, have not been examined before; that is agreeing to disagree. Although there have been increasing concerns about the polarization of views and echo chambers in online spaces (Freelon, 2015; Veletsianos et al., 2018; Walter et al., 2018), the fact that some users *agree to disagree* shows that, despite remaining entrenched with opposing stances, it is possible for users to engage in intersubjectivity or at least take on board others' stances. Altogether, the findings in this chapter contribute to the growing literature that shows that disagreement can be constructive, if users employ discourse practices that are facilitative of intersubjectivity and exploration of alternative voices (Baym, 1996; Bou-Franch & Garcés-Conejos Blitvich, 2014; Chiu, 2008; Concannon & Healey, 2015; Felton et al., 2015; Marra, 2012; Mercer, 2004).

It is worth relating the findings in this chapter to MOOC research. The general patterns of replies and the length of threads suggested short-lived interactions, which can be an indication of cumulative talk and information exchange, as has been found by Poquet et al. (2018) and Wise et al. (2016) in other MOOCs. What is new to the MOOC research is that the micro-analysis illustrated that users with opposing stances engage in online deliberation, although they may remain polarized. Similar to Drasovean & Tagg's (2015) findings on TED, users' discourse practices for raising disagreement can also be maintaining relationship at the same time, further suggesting that categorizing a comment as performing a single function such as disagreement, or question is rather incomplete. Furthermore, the topics users discuss are not necessarily dictated by course design and may not be considered as on topic by Wise et al. (2016), yet suggesting the opportunity of MOOCs as a third space (Wright, 2012). This will be further discussed in Chapter 10.

Finally, an unexpected reply keyword *link* points to the URL-posting practices in online spaces. The next chapter delves into this particular keyword and reveals that URLs are also employed when users disagree with each other, and can at times become a barrier to the processes of intersubjectivity.

Chapter 9

Discourse practices of URL-posting

9.1 Introduction

Unlike the previous chapters that looked at discourse practices in the initiating posts, independent posts and replies, this chapter zones into one discourse practice specific to online spaces, URL-posting, and addresses the third research question:

RQ3: How does URL-posting initiate, sustain or hinder dialogic conversations in online discussions?

Informed by previous studies that show that users employ various online sources linked to URLs as evidence for stance-taking (Jacobson et al., 2016; Savolainen, 2014; Wikgren, 2003), and the reply keyword *link* found in Chapter 8, this chapter explores *how* users employ URLs and respond to URLs, especially when they hold opposite stances and engage in sustained interactions. Therefore, to some extent, this chapter also continues Chapter 8's exploration of users' discourse practices for negotiating their disagreement. Specifically, the analysis highlights the potential problems of *link wars*, in which both or either parties in disagreement present URLs to support their stances.

Specifically, this chapter conducts micro-analysis to investigate how users with different stances engage with each other by employing URLs, and how others respond to them, extending previous studies that thus far have only examined limited discursive context of URL-posting (Polletta et al., 2009). The analysis thus addresses RQ3 regarding how URLs can facilitate or hinder dialogic conversations in online discussions. Given that this analysis focus on users' discourse practices around URLs, the findings also provides insights into users' co-construction of the meaning of URLs in online spaces.

This chapter begins with a quantitative analysis of the extent of URL-posting in FutureLearn MOOCs. To understand the general discourse patterns in URL-posting, a collocation analysis of URL addresses posted and the reply keyword *link(s)*, and an analysis of the length of the comments containing the URLs, are conducted. A series of micro-analyses of threads in which users employ URLs in their stance-taking is then conducted to reveal the discourse practices of URL-posting in online discussions.

9.2 Extent of URL-posting on FutureLearn

The number of URLs posted in each course, different types of comments, and individual differences in including URLs in their comments are first examined to describe the extent of URL-posting in the FutureLearn online discussions. Next, the sources linked to the URLs are categorized to examine the online information sources employed by users.

9.2.1 URL-posting varies across courses and types of comments

The number of URLs posted in each type of comment in each course (see Table 9.1) are calculated based on the occurrence of search strings `http.+|www.+`. URLs per comment is used for comparison across different courses and comments. A total of 8813 URLs are found in 7243 out of 202787 users' comments, equivalent to 0.04 URL per comment or 1 in every 2500 posts, consistent with what have been found in previous studies that URLs are one of the many sources and contents that users refer to in online discussions (Oh et al., 2008; Polletta et al., 2009; Wikgren, 2001).

Table 9.1 URLs posted in each type of comment in each course

	Initiating Posts		Independent Posts		Replies ¹		Total	
	Number of URLs	URLs per comment	Number of URLs	URLs per comment	Number of URLs	URLs per comment	Number of URLs	URLs per comment
accessibility-2	153	0.13	228	0.05	122	0.06	503	0.06
ancient-1	808	0.19	1239	0.14	832	0.08	2879	0.12
code-1	138	0.07	460	0.10	241	0.07	839	0.08
corpus-1	236	0.04	79	0.02	285	0.06	600	0.04
dyslexia-1	246	0.05	336	0.01	176	0.03	758	0.02
finance-1	59	0.03	73	0.03	183	0.03	315	0.03
management-4	16	0.01	11	0.00	17	0.01	44	0.01
moons-1	219	0.05	232	0.02	452	0.06	903	0.04
nutrition-4	146	0.05	267	0.01	195	0.04	608	0.02
oceans-1	141	0.09	239	0.06	106	0.06	486	0.06
palliative-1	10	0.01	44	0.01	19	0.01	73	0.01
soils-1	154	0.08	221	0.03	430	0.11	805	0.06
Total	2326	0.07	3429	0.03	3058	0.06	8813	0.04

¹ First contributions and subsequent contributions are collapsed into one category of replies due to their small number.

Perhaps due to the learning design of each course, URL-posting varies across the courses.

Ancient-1 (0.12 URL per comment) and code-1 (0.08 URL per comment) have the highest URL-posting rate, whereas management-4 and palliative-1 (0.01 URL per comment) have the lowest URL-posting rate. In ancient-1, users are asked to look for information online, for example “Go online and find information about the diet of the ancient athlete Milo of Croton, and of other Pythagoreans. You will find conflicting stories about what Milo ate. What did you find out about the way ancient Greeks or Romans viewed vegetarianism?”¹⁴³. In code-1, at the start of the course, users are encouraged to share their projects as they go through the course, “You can then share the single PDF or HTML file as before, by email, via Dropbox or Google Drive, on your blog and via a link in the discussion below.”¹⁴⁴ In contrast, in management-1 and palliative-1, users are typically asked to share their work experience in contract management or palliative care, for example “What barriers

¹⁴³ <https://www.futurelearn.com/courses/ancient-health/1/steps/154833>

¹⁴⁴ <https://www.futurelearn.com/courses/learn-to-code/1/steps/52679>

to change have you encountered? How were they resolved?”¹⁴⁵, “Have you been surprised by what you have heard so far? How does it compare with your experiences in your own country?”¹⁴⁶ This contrast between sharing URL and personal experience will be further explored in section 9.7. where users participating in the same thread use these two types of evidence for stance-taking.

The relationship between learning design and URL-posting is also consistently found in other courses when there is a prompt asking users to share resources. For example, in accessibility-2, “Please share any resources you might know about that would help those with communication difficulties in different countries as English (spoken and written) often dominates and yet communication happens in many different ways being multilingual and multicultural.”¹⁴⁷ and in dyslexia-1, “Are there any specific tools, applications, software, websites that you would recommend?”¹⁴⁸

The URL-posting prompted by this learning design is information sharing in nature as users respond to the prompt that invites them to share resources. However, this does not necessarily mean that URL-posting is always user-content practice which appears only in independent posts. The distribution of URLs in the initiating posts (0.07 URL per comment) and replies (0.06 URL per comment) is relatively more frequent than in the independent post (0.03 URL per comment), suggesting that URL-posting might have the potential to invite and sustain user-user interactions. This finding is consistent with Gallagher & Savage's (2016) work on another FutureLearn MOOC. However, this quantitative analysis only shows trends but not significant differences. More importantly, it does not explain how URL-postings in different types of comments can enable or hinder user-user interactions with others in the discussions. Thus, how users use URLs in their comments will be explored later in the micro-analysis in section 9.5.

¹⁴⁵ <https://www.futurelearn.com/courses/contract-management/4/steps/115187>

¹⁴⁶ <https://www.futurelearn.com/courses/palliative/1/steps/97015>

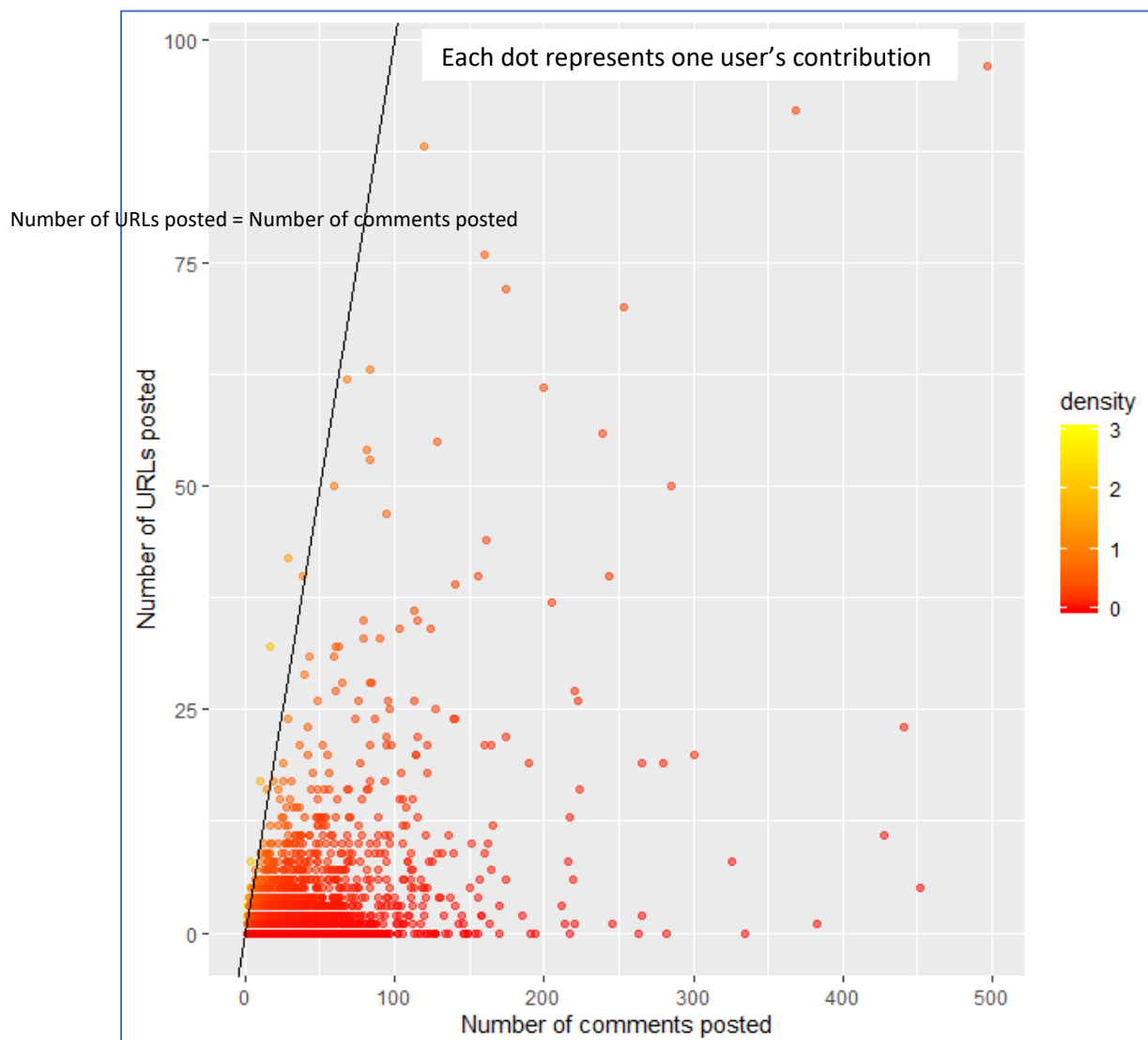
¹⁴⁷ <https://www.futurelearn.com/courses/digital-accessibility/2/comments/19541309>

¹⁴⁸ <https://www.futurelearn.com/courses/dyslexia/1/steps/22664>

9.2.2 URL-posting varies across individuals

URL-posting also varies across individuals, as shown in Figure 9.1. Among 23495 users who post at least one comment in the online discussions, 89% never includes an URL. Among the rest of the 2413 users who post URLs, the number of URLs posted ranges from 1 URL to 97 URLs, and from 0.003 URL per comment to 3 URLs per comment. This variation parallels the distribution of sources used in Q&A forums where some users rely on internet resources, i.e., URLs, in their postings, while others rely on other kinds of information and evidence, including their own knowledge, personal and others' experiences (Oh et al., 2008).

Figure 9.1 Scatterplot: Number of URLs posted vs. Number of comments posted.



Note. Each dot represents a user's posting behaviour in terms of number of comments and URLs posted. The reference line is where the number of URL posted equals to the number of comments posted. The density refers to URLs per comment.

Eighty-five out of the 120 super-posters defined in this thesis post at least one URL and together contribute 15% of the 8813 URLs found in the corpus. Among these super-posters, 16 of them have posted 0.2 URLs per comment, translating to one fifth of their comments, suggesting that some super-posters post URLs particularly frequently in the online discussions. Furthermore, 54 users who contribute at least 10 comments in the discussions post 0.5 URLs per comments, translating to half of their comments, suggesting that some users rely heavily on URLs. This finding not only raises the question why some users include so many URLs in their comments, but also how they make use of them, especially when they interact with those who do not rely on URLs. This

question will be addressed by investigating users' discourse practices of posting URLs and meta-pragmatic discussions on URLs in section 9.5.

9.2.3 Major sources of URLs posted

Table 9.2 presents the major categories of sources linked to the URLs that users post in FutureLearn discussions. The categorization only accounts for 45% of all the URLs posted. This is because, due to the large number of URLs, only those URLs with top-level domain names or paths mentioned at least 10 times are examined. This categorization strategy follows Oh et al. (2008) and Sudau et al. (2014).

Table 9.2 Major sources of URL posted.

Sources	Descriptions	Frequency ¹	Percentage
News Media	News, news aggregators, or magazine websites	736	8%
Higher Education	Institutions of higher education, measured based on top-level domain .edu and .ac	646	7%
Video	YouTube, Vimeo, Ted	590	7%
Governmental and International Institutions	Government establishment and international treaty-based organisations, measured based on top-level domain names including .gov and .int. NHS is also included here	531	6%
Wikipedia	Wikipedia, Wikimedia	523	6%
File Sharing and Hosting	Github, Google Drive, Onedrive, Dropbox	330	4%
Blogs	Wordpress and Blogspot as well as URLs with blog or blogs in the path	189	2%
Academic Journals	Jstor, Science Direct, OUP, Nature, Wiley, and Researchgate and Academia	143	2%
Social Media	Flicker, Pinterest, Facebook, Google Plus, Twitter, Linkedin	137	2%
Books	Books on Amazon, Google Book, Goodreads, itunes, Project Gutenberg	124	1%

¹This is a conservative count because only the domain name with at least 10 mentions are categorized. There should be more sources in each of these categories.

The sources posted by users include websites by established news media, governmental and educational institutions, books, and social media, as in other online discussions (Oh et al., 2008; Wikgren, 2003). Commercial websites, which are accounted for by previous studies, are not identified here because it requires thorough examination of the websites. As a category, news media is relied on the most, probably because in online learning settings, conventional news media are perceived as more credible, compared to commercial information and social media. Also, news media is perhaps shorter and easier to read compared to academic journal articles. However, in my concordance reading, I have also come across users expressing their doubt towards news media linked to (see section 9.5 for examples). It is worth noticing that, although fewer than other sources, academic sources are also one of the sources users rely on probably because of the academic nature of MOOCs. Lastly, sometimes users introduce their own websites and organizations they support.

YouTube and Wikipedia are the two most linked-to websites, as expected for their popularity (Khan, 2017; Singer et al., 2017). These two websites are typically used for information sharing (Khan, 2017; Oh et al., 2008). YouTube content varies from professional-generated to user-generated contents, such that some YouTube URLs in the FutureLearn corpus may link to content attributed to news media, higher education or blogs (Welbourne & Grant, 2016), although the current study does not investigate this aspect. Besides information sharing, users might rely on YouTube and Wikipedia for an overview of the subjects for their own information seeking (Khan, 2017; Singer et al., 2017). Similar to news media, users also have doubt over Wikipedia, as evidenced in their discourse, which will be explored in section 9.5.

Based on the major sources of URLs posted, as shown in Table 9.2, it can be argued that users rely on two key types of sources: 1) established sources, such as governmental, international and academic institutions, academic journals, and news media; 2) users-generated sources, such as videos, social media, blogs, and Wikipedia. The established sources account for 52% of all the sources categorized in the current study, whereas the user-generated sources 36%. However there may be more user-generated sources, as well as commercial websites, in the corpus because other

URLs with .com domain are not categorized. This division is consistent with Connor (2013), Sudau et al. (2014), Savolainen (2014), Wikgren (2003) who identify individual differences in users' reliance on established and authoritative information contained in URLs, versus personal experience, popular media and persuasive materials contained in URLs. Although URLs are less used compared to personal and situational experience, the expertise or authoritative information as presented in the URLs posted is typically valued more by users (Koschack et al., 2015; Oh et al., 2008; Polletta et al., 2009; Savolainen, 2014; Shanahan, 2010). The findings point to the possibility that users relying on different types of sources may come together in the online discussions and may have disagreement on each other's use of sources for evidencing their stances. This difference in evidencing practices will be explored in section 9.7.

9.3 General discourse patterns of URL-posting on FutureLearn

The general discourse patterns of URL-posting on FutureLearn are examined with a collocation analysis and the length of comments containing URLs. Collocation analysis reveals words that co-occur frequently with the word of interest, i.e., the URL addresses¹⁴⁹ and the reply keyword *link(s)* in the current analysis. The descriptive analysis of length of comments containing URLs reveals whether users simply just post URLs with few or no own words in their comments.

9.3.1 Collocation analysis

9.3.1.1 Collocates of URLs: how users include URLs in comments

The collocation analysis of the URL addresses posted reveals three main ways by which users include URLs in their comments – the URLs are explicitly introduced or referred to, included as an in-text citation, or posted as if it is a reference in academic writing. These patterns are similar to other online spaces (Polletta et al., 2009; Wikgren, 2003), and are identified by collocation analysis and

¹⁴⁹ The regular expression used in the corpus tool is `http.+|www.+`

concordance reading, rather than strict categorization. There might be instances which could fall into more than one category.

URLs are explicitly introduced or referred to

The collocates found to be occurring within a 5-word window to the left of URLs posted in users' comments show that users typically introduce URLs by signposting with *here, this, example, following*, and referring to URLs with meta-language, such as *link, article, source, website, video, page*. Verbs such as *see, read, try, look* are used to invite others to visit the URLs. Positive evaluations such as *interesting, useful* are also used when users introduce their URLs. The use of these collocates indicate that users introduce and refer to the URL explicitly. One example containing some of the collocates is as follows:

[1] *Interesting video link here*. Even if you don't believe the link shows all that depends on soil <http://www.andiesisle.com/creation/magnificent.html>"¹⁵⁰

This discourse of explicitly introducing or referring to the URLs posted can be an invitation to others to have a look at them (Polletta et al., 2009). It is likely to be an act of information sharing, although there can be other purposes. For example, the explicit introduction, "Interesting video link here" in [1] is accompanied by a persuasive discourse, "if you don't believe the link shows all that depends on soil", suggesting the URL is also used as an evidence for the claim "all that depends on soil". However, the user does not mention exactly why. When the URLs are explicitly introduced or referred to in the comments, users sometimes only write minimally, such that the URLs become the main substance in the comment, as shown in example [1]. Other users might have to visit the URLs to understand what the comments are about or why the URLs support the claim.

URLs are included as an in-text citation

¹⁵⁰ <https://www.futurelearn.com/courses/soils/1/comments/6225078>

Other collocates indicate that URLs are used in a way similar to the in-text citations in academic writing. This is usually in comments where users write in length to support their stance-taking. The URLs posted are where the users draw information, credentials, evidence or argument from for their comments, and readers do not necessarily need to visit the page to understand what has been written (Myers, 2009). One common means of citation in the online discussions is to put the URL address in a parenthesis, as shown in the example below:

[2] “.....It appears to me that the American guidelines settle at about half the UK amount of omega-3 fatty acids (<http://www.seafoodnutrition.org/blog/2015-dietary-guidelines-for-americans-announced>) and if both groups work from the same data, I wonder why the difference.”¹⁵¹

[3]“.....There are numerous organisations that could support this. For example, I use to be a trustee for Charities Evaluation Services (<http://www.ces-vol.org.uk/tools-and-resources>) and they doing amazing work supporting charities of sizes of around 10-100 employees improve governance, transparency and recording outcomes.....”¹⁵²

In [2], the URL is used to indicate the evidence supporting the user’s claim that “the American guidelines settle at about half the UK amount of omega-3 fatty acids”. In [3], the URL is to evince the existence of “Charities Evaluation Service” which is an “example” of the “numerous organisations” in the user’s claim.

Sometimes users also cite the URL in text, as shown in the following examples:

¹⁵¹ <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/19264729>

¹⁵² <https://www.futurelearn.com/courses/exploring-our-oceans/1/comments/483626>

[4] “.....According to <http://www.universetoday.com/102630/how-many-stars-are-there-in-the-universe/> there is a septillion stars, so I'll have to give my vote to more microbes. The question was about the observable universe, beyond that no one knows.”¹⁵³

[5] “.....see the attached link (it is in the guardian but the figures quoted are DWP) <http://www.theguardian.com/news/datablog/2013/jan/08/uk-benefit-welfare-spending> This shows that if you add in over 75s TV license , benefits to pension support, disability allowance carers allowance etc that an awful lot of benefits are aimed at pensioners or other people who cannot work.....”¹⁵⁴

In [4], the URL is indicated as source of evidence with the phrase “According to” to support the user’s stance “there is a septillion stars, so I'll have to give my vote to more microbes”. In [5], although the URL is explicitly introduced with “see the attached link”, the phrase “this shows” indicates that the URL is an evidence that supports the user’s claim that follows, “if you add in over 75s TV license.....”. The in-text citation of URLs typically occurs in comments where users write in length about their stance and how the content of the URLs relates to it, such that visiting URLs might be optional.

URLs as references

URLs sometimes are presented at the end of the comments or follow the collocate *source*. This is comparable to the reference list in academic publications where readers can find the original source that a user’s claim is based on. Similar to in-text citation, this typically occurs in comments where users write extensively in their own words. One example is presented below:

[6] “Sustainable living is a lifestyle that attempts to reduce an individual's or society's use of the earth's natural resources and personal resources. Practitioners of sustainable living often

¹⁵³ <https://www.futurelearn.com/courses/exploring-our-oceans/1/comments/362180>

¹⁵⁴ <https://www.futurelearn.com/courses/inequalities-in-personal-finance/1/comments/4579175>

attempt to reduce their carbon footprint by altering methods of transportation, energy consumption and diet.....

Source : www.biologicaldiversity.org , www.wikipedia.org¹⁵⁵

In short, the collocation analysis of the URLs posted in the online discussions reveal three general means of including URLs in one's comment. Admittedly, they are not very clear cut. As shown in [5], it can be an explicit introduction of the URL ("see the attached link") but also a citation of evidence for one's claim in the comment ("This shows that"). Based on users' wording surrounding the URLs posted, it can be deduced if the URLs are posted for information sharing, or for supporting one's stance. Typically, an URL is posted for information sharing when the comment only contain minimal wording that explicitly introduce the URL (Polletta et al., 2009). When an URL is posted like a citation or reference, users include them for supporting their stance-taking (Wikgren, 2003). Therefore, there are three means of including URLs in comments, which serve two key functions, information sharing and stance-taking, similar to what have been observed in other online spaces (Connor, 2013; Jacobson et al., 2016; Polletta et al., 2009; Savolainen, 2014; Wikgren, 2001).

9.3.1.2 Collocates of *link(s)*: how users respond to URLs

The collocates of *link*, and its plural form *links* in the users' replies are examined because *link* is a reply keyword, as revealed in Chapter 8. Importantly, users' replies are more likely to reveal their reactions towards URLs posted in the threads, because replies are in response to initiating posts and others' replies within a thread. The collocates found include positive evaluation, such as *great*, *interesting*, *good*, words of appreciation, *thank(s)*, and verbs such as *sharing* and *posting*, suggesting that users acknowledge and value the URLs posted, as shown in the replies in Figure 9.2. Previous studies also observe similar responses to URLs posted (Polletta et al., 2009; Wikgren, 2001).

¹⁵⁵ <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/19188496>

Figure 9.2 Thread 18971576.

Thread 18971576
Source: <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/18971576>

Initiating Post 2017-01-19 10:12:40 Likes: 4
User n4-2223

The Japanese, known for their longevity, [.....]. The Chinese use a Food pagoda [.....] Both these and other historical permutations of the pyramid are shown here <http://discovermagazine.com/galleries/zen-photo/f/food-guides> and it's interesting to compare the slight differences between cultures and eras. I wonder do people from different countries need different portions when you think about differences in body sizes and food tolerances (many Japanese cannot tolerate lactose for example). These food guides can never be more than a generalisation as surely the needs of each individual also vary depending on the body's condition at any given time. [.....]

Reply 1 (first contribution) 2017-01-19 14:21:37 Likes: 1
User n4-3469
Thank you, very good link.

Reply 2 (first contribution) 2017-01-20 00:45:34 Likes: 0
User n4-217
it is a great link, thanks.

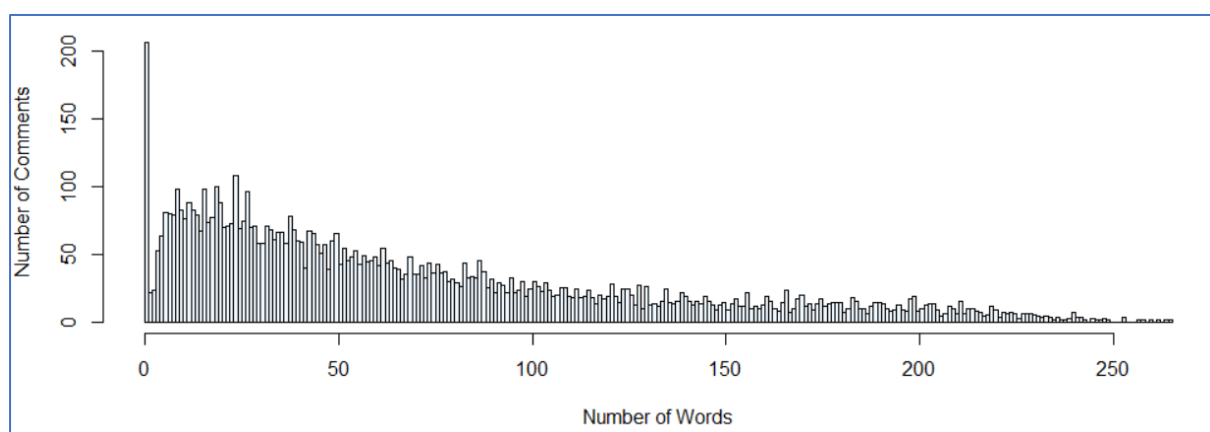
This collocation analysis speaks to the information-sharing function of URLs in online discussions, as shown by other users' acknowledgement that the URLs as "great/good link". However, in Figure 9.2, in the initiating post containing the URL, the user seems to include the URL as an in-text citation, "Both these...are shown here" and elaborates on their stance following the presentation of the URL, "I wonder do people from different countries need different portions". This stance is not taken up by either of the replies which only acknowledge the URL. A discussion is thus not generated by the initiator's stance-taking, perhaps due to other users' focus on the URL. This example suggests that the URL is not only used for information sharing but to take a stance, although it can be subjected to different interpretations by other users, who generally express positive sentiments towards the URLs posted.

9.3.2 Length of comments containing URLs

The number of words in the 7243 comments containing URLs is calculated to examine how much users write in a comment besides the URLs (Figure 9.3), thus revealing their discourse practices of including URLs in their comments. When users only post URLs without their own words, the wordcount is zero. Twenty-five percent of the comments containing URLs are within the word count of 23, of which 198 are zero wordcount, as can be seen in the spike in Figure 9.3. These 198 URLs are termed as *unaccompanied URLs*, because the comment consists only of URL(s). Among them, 109 are independent posts, 22 initiating posts, 67 replies. The observation that there are some initiating posts and replies containing only unaccompanied URLs attests to the possibility that the URL itself may have communicative functions for user-user interactions in threads, not to mention when users include URLs in longer comments.

In the following subsections, examples of unaccompanied URLs, and comments of different lengths containing URLs are presented to illustrate how users include URLs in their comments to achieve their communicative functions. Besides confirming the general functions of URL-posting as information sharing and stance-taking, it is also found that sometimes users substitute their own voice with the URLs, i.e., they do not write their stance in their comments but defer to the URLs to do the job for them. In this case, in order to understand them, the other users would have to visit the URLs.

Figure 9.3 Number of words in the comments containing URLs.



9.3.2.1 Unaccompanied URLs

Unaccompanied URLs are posted without any words from the users. It can be a way of sharing information and substituting one's voice with information linked to the URLs, since it is rather convenient for users to copy and paste the URLs, and they may not have additional comments on top of the content linked to the URLs. These two possible communicative functions can only be deduced from the context where the unaccompanied URLs are posted. Three examples of unaccompanied URL posted in independent post (Figure 9.4), initiating post (Figure 9.5) and reply (Figure 9.6) are illustrated below.

Figure 9.4 Unaccompanied URL in an independent post

Thread 19204328

Source: <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/19204328>

Discussion Prompt:

Each country has its own set of dietary guidelines to give reference values to promote healthy eating.

What are the dietary reference values in your own country? Are there other visual representations, like the 'eatwell plate', that can be used to inform healthy eating?

Independent Post 2017-01-29 00:16:17 Likes: 0

User n4-1896

<https://health.gov/dietaryguidelines/>

Figure 9.5 Unaccompanied URL in an initiating post

Thread 19479791
Source: <https://www.futurelearn.com/courses/digital-accessibility/2/comments/19479791>

Discussion Prompt:
Exercise: try out a screen reader
How did you get on? Was it easy to use? Did you encounter content that was difficult to access?
Were there unexpected difficulties?

Initiating Post 2017-02-07 Likes: 0
User a2-530
<https://www.quora.com/Do-blind-children-really-need-to-learn-brail>

Reply 1 (first contribution) 2017-02-08 Likes: 1
User a2-589
Thanks for sharing [Learner a2-530]. Braille is approaching 200 years. A quick enquiry on google chrome "the future of braille came back with 543,000 hits. A second query "does braille have a future" - came back with 5.7 million hits. A third query "will braille become redundant" came back with around 430,000 hits. Also so many articles, reports etc. on the theme of braille under threat.

Reply 2 (first contribution) 2017-02-22 Likes: 0
Facilitator
Here's an article from 2006: "Proponents Say the Decline in Braille Instruction Is Leading to Illiteracy": <https://nfb.org/images/nfb/publications/bm/bm06/bm0609/bm060905.htm>.

In both Figure 9.4 and Figure 9.5, unaccompanied URL is posted as a new post that is in response to a discussion prompt. In the independent post in Figure 9.4, the unaccompanied URL links to a health guideline, which is what the discussion prompt asks for, "What are the dietary reference values in your own country? Are there other visual representations ...". Thus, the user seems to use the URL to substitute their voice in response to the prompt, and also to refer intertextually to images that is not possible to post in the online discussions (Kiernan, 2018). It is also an act of information sharing via the URL, given that the discussion prompt is asking for information regarding users' "own country". Because this unaccompanied URL receives no reply or like, it is hard to determine other users' interpretation of the URL.

In the initiating post in Figure 9.5, the unaccompanied URL posted is linked to the Quora discussion forum on the question of “Do blind children really need to learn brail”, as indicated by the URL address. Links to other forum messages are also found in other online spaces, and the sources are considered as personal opinions (Connor, 2013; Polletta et al., 2009; Savolainen, 2014; Sudau et al., 2014). In this case, the URL posted is not strictly related to the discussion prompt that asks users to share their reflection, “How did you get on?” after using “a screen-reader”, but related to the general topic of blindness. Although it is not in response directly to the discussion prompt, this unaccompanied URL receives a reply “thanks ... for sharing”, suggesting that other users interpret this unaccompanied URL as information sharing and express gratitude for it. The unaccompanied URL in this initiating post also forms a common ground for others to continue the thread, as evidenced by another user’s further addition about their own “query on google” in reply 1 and one facilitator’s reply with another URL with minimal wording “here’s an article”. URL-posting may thus trigger cumulative sharing of more information via URLs, similar to the cumulative talk conceptualized by Mercer (2004).

Additionally, the unaccompanied URL in the initiating post in Figure 9.5 may be considered by other users as reflecting the user’s stance-taking towards the content linked to the URL. The URL posted is on “Do blind children really need to learn brail”, which is aligned by reply 1 “braille under threat” and the URL in reply 2 “Decline in Braille”. These two replies suggest other users may have considered this unaccompanied URL as stance-taking such that they also provide similar information to align. Thus, unaccompanied URLs can also be used to represent one’s voice in stance-taking, although the user does not write explicitly their own stance. Using URL to represent one’s voice in stance-taking can also be found in replies, as shown in Figure 9.6.

Figure 9.6 An unaccompanied URL in reply 8

Thread 6349382

Source: <https://www.futurelearn.com/courses/soils/1/comments/6349382>

Initiating Post 2015-07-21 09:49:14 Likes: 1

User s1-310

Very interesting, but the Wensum area is in East Anglia not the South East .

Reply 1 (first contribution) 2015-07-21 11:11:09 Likes: 0

User s1-1518

Hi [s1-310] - and the Avon shown on the map is not in the south west - have they marked the wrong Avon?

Reply 2 (first contribution) 2015-07-21 21:11:47 Likes: 0

User s1-1260

Probably. Avon means river so there are several River Avons. There's one that goes through Bristol and one in Devon, which rises on Dartmoor and flows south.

Reply 3 (first contribution) 2015-07-24 19:34:37 Likes: 1

User s1-981

From Lancaster I think it is reasonable to lump East Anglia as part of the SE and similarly where they classify the SW. I am from London and consider Birmingham to be the start of the North. Lol :)))

Reply 4 (subsequent contribution) 2015-07-25 06:33:33 Likes: 0

User s1-1518

Hi [s1-981] - coming from Coventry I prefer to think of the Midlands as the Heartland and everywhere else as the periphery!

Reply 5 (first contribution) 2015-07-26 11:49:07 Likes: 0

User s1-599

Was she not referring to the WANTSUM channel which is indeed in the south east being almost as far SE as you can get in Kent. Since AVON was Anglo Saxon for River there are indeed several rivers called 'Avon'

Reply 6 (initiator's subsequent reply) 2015-07-26 16:34:58 Likes: 1

User s1-310

Noooo, East Anglia is not in the South east it's in the east East !

Reply 7 (subsequent reply) 2015-07-27 07:47:18 Likes: 1

s1-599

Looking back at the map, I was sewing while listening, I realise she did mean Wensum and you are of course correct it is the East. Wonder if her degree was in Geography?

Reply 8 (first contribution) 2015-07-27 15:46:31 Likes: 0

User s1-1350

All

<https://localreviews1.knoji.com/the-eight-river-avons-of-britain/>

Reply 9 (subsequent contribution) 2015-07-28 07:10:32 Likes: 0

User s1-599

Thank you [s1-1350] I stand corrected.

The thread in Figure 9.6 is a debate about some location issues, as can be seen from the initiating post ('the Wensum area is in East Anglia not the South East') to reply 7 ('you are of course correct it is the East'). Amid this debate, a user replies with an URL while addressing the other users with "All" in reply 8. Although not strictly an unaccompanied URL, it is without any other user's own wording. This URL can be interpreted as both sharing information and taking a stance in the debate, as evidenced by the response in reply 9 "Thank you" and "I stand corrected". It is possible that the previous replies in the thread may have already given the context of the URL, such that the URL-posting user do not need to write about the URL. These three examples illustrate that unaccompanied URLs may have the communicative function of sharing information and representing ones' voice in the online discussions.

9.3.2.2 Minimal wording with the URLs

In contrast to unaccompanied URLs, there are 1628 comments containing URLs but with *minimal wording* of the users. In this thesis, I define them as those with wordcount fewer than or equal to 23. The cut-off is arbitrary but based on statistics. These comments with minimal wording, along with those unaccompanied URLs comprise the shortest 25% comments that contain URLs. This compares to the 33% of links introduced with minimal wording in Polletta et al.'s (2009) study, although they did not report how long the comments are.

There are 667 independent posts, 307 initiating posts, and 654 replies containing URLs with minimal wording. The minimal wording mainly consists of collocates which are used to explicitly introduce or refer to the URLs posted, as documented in section 9.2.1. For example, "If you are interested in the Moon check this out: <http://lroc.sese.asu.edu/images/videos>"¹⁵⁶, in which "check this out" is used to explicitly introduce the URL. Without a longer elaboration or explanation of the URLs, it might be hard to deduce how users are applying information from the URLs. In this example of minimal wording, there is no indication of them citing the URL for supporting claims. This URL-

¹⁵⁶ <https://www.futurelearn.com/courses/moons/1/comments/828699>

posting may be more an act of information sharing, as evidenced in the introductory frame, “If you are interested”. This introductory frame can also be a way of establishing a dialogue with potential users, as revealed in Chapter 6.

Within the minimal wording, brief mention about the main content of the URLs is another way users introduce URLs in their comments. For example, “Interesting article with subduction and water in mind ...<http://www.scientificamerican.com/article/rare-diamond-confirms-that-earths-mantle-holds-an-oceans-worth-of-water/>”¹⁵⁷ where “subduction and water” is mentioned as the main point of the content linked to the URL. Although these URL-containing comments do not necessarily receive replies from others, as attested by the number of independent posts, the language use in these brief introductions of the URLs establish an invitation to other users to visit the URLs posted, pointing to the information sharing function of URL-posting (Polletta et al., 2009).

As with the unaccompanied URLs, there are times that URLs are used to substitute users’ voice in comments with minimal wording. This is explicitly indicated by the user’s minimal wording, as illustrated by the initiating post and reply 2 in Figure 9.7.

¹⁵⁷ <https://www.futurelearn.com/courses/exploring-our-oceans/1/comments/492601>

Figure 9.7 Minimal introduction of URLs in the initiating post and reply 2.

Thread 18936774
Source: <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/18936774>

Initiating Post 2017-01-17 23:18:04 Likes: 3
User n4-227
There is serious debate about whether the "calories in=calories out" theory is valid.
<http://harvardmagazine.com/2016/05/are-all-calories-equal>
<https://authoritynutrition.com/debunking-the-calorie-myth/>

Reply 1 First reply 2017-01-18 03:02:48 Likes: 0
User n4-3719
YES!!!

Reply 2 First reply 2017-01-18 09:51:36 Likes: 1
User n4-594
Hi [n4-227], thanks for the interesting articles! I want to contribute to the discussion with some other reading resources also:
<https://examine.com/nutrition/high-carb-high-satiety/>
<http://weightology.net/gary-taubes/good-calories-bad-calories-the-mythology-of-obesity-or-the-mythology-of-gary-taubes.html/>
<http://georgiefear.com/2016/12/15/the-truth-about-ketogenic-dieting/>
<http://www.vox.com/2016/7/6/12105660/do-low-carb-diets-work>
<http://www.trainerize.com/blog/a-beginners-guide-to-eating-for-weight-loss/>
Stay well!

Reply 3 Initiator's reply 2017-01-18 19:59:14 Likes: 0
User n4-227
A lot to digest here. (pun intended) i'm having problems opening one or two of your links. Will try again later. Always good to read the opinions of others.

In the initiating post in Figure 9.7, the user introduces two URLs by questioning the validity of the “ ‘calories in=calories out’ theory” with minimal wording. This wording gives a brief information regarding the content linked to the URLs and could be interpreted as information sharing, rather than stance-taking. However, one user reply with a word “YES” in reply 1, which suggests an alignment with what has been posted. Thus, the URLs posted seem to be interpreted as taking stance against the “ ‘calories in=calories out’ theory”. In reply 2, another user acknowledges the URLs posted with “thanks for the interesting articles!”, suggesting that the URL-posting in the initiating post is interpreted by this user as information sharing. In reply 2, this user also introduces

other URLs with minimal wordings “some other reading resources”, suggesting information-sharing. This information-sharing is indeed acknowledged by the initiator with “always good to read the opinions of others” in reply 4. The user contributing more URLs in their reply following the posting of URL in the initiating post is similar to the cumulative sharing of URLs, as observed in Figure 9.5.

Thus, the URL-posting in both initiating post and reply 2 in Figure 9.7, which only consist of users’ minimal writing, can be deemed as information sharing. However, interestingly, user n4-227 seems to suggest that they are engaging in a discussion, as evidenced in reply 2 “I want to contribute to the discussion”. But, they do not really write their own stances but mainly present URLs, as shown in both the initiating post and reply 2. Therefore, it might be hard for other users to really understand this “debate” or “discussion” unless they visit the URLs. This attests to the previous observations on unaccompanied URLs that users employ URLs to substitute their voices in stance-taking, instead of writing their own.

Posting URLs without one’s own words, as in unaccompanied URLs, or with minimal wording reveals little to no direct information regarding the content linked to the URLs. Nonetheless, the analysis thus far shows that these URLs may also be employed by users to substitute their voice. This is especially possible when they respond to the discussion prompt or put an URL in the middle of a thread where others have been debating. Given the technology affordance of online spaces, it is rather convenient for users just copy and paste the URLs rather than typing out in length. This practice of URL-posting may mirror the URL-sharing on social media where users seldom write about what they share, but at the same time use it as a way to express themselves (Edgerly et al., 2016; Oeldorf-Hirsch & Sundar, 2015).

Nonetheless, the lack of users’ own words requires the agency of other users to read and interpret the function of the URLs posted (Colaric & Jonassen, 2001). Although the collocation analysis shows that most users are positive towards the URLs posted, the posted URLs could be interpreted differently, as shown in the examples above. This could become a problem when it is a contentious issue, which I will come back to in section 9.5. This minimal to no elaboration of the

URLs posted also contrasts with the in-text citation or referencing, as shown in section 9.3.1, where the URLs are posted as the source of supporting evidence for users' stance that they write about in length, and readers do not necessarily need to visit the cited URLs to make sense of what has been written (Myers, 2009; Wikgren, 2003).

9.3.2.3 Long comments incorporating URLs

Seventy-five percent of comments containing URLs consists of users' own wordings ranging from 24 to 265 words (*long comments*). Users could be writing in length about the URLs, or their stances while using the URLs to support their stances, although some users might be quoting verbatim from the websites linked to the URLs. These various scenarios are illustrated in the following examples.

Introducing URLs with own reflections about the URLs

In contrast to the minimal wording, some users write in length about the URLs they post, while explicitly introducing or referring to the URLs. This allows them to not only share the information linked to the URLs but also write their reflection. This is illustrated in the initiating post in Figure 9.8 and reply 1 in Figure 9.9.

Figure 9.8 The URL is introduced in length in the initiating post.

Thread 6283456

Source: <https://www.futurelearn.com/courses/soils/1/comments/6283456>

**Initiating post 2015-07-14 20:24:25 Likes: 1
s1-1966**

Here is a great piece of research from the UEA on the antibiotic properties of *Streptomyces* sp. bacteria as found in leaf-cutter ant colonies: <https://www.uea.ac.uk/events/whats-on/event-recordings> Apparently, warmed soils and mouldy breads were once used to prevent infections, an ancient remedy that worked due to the same antibiotic properties...!

**Reply 1 (first contribution) 2015-07-19 19:52:27 Likes: 1
s1-1417**

Thanks very much for the link! Being an MD I am absolutely astonished. The soil seems to harbour even new strategies to treat infections! There are much more aspects about soil I've ever imagined. I was surprised that even human pathogens like *pseudomonas* are able to live in soil. I have guessed that they were only able to survive at body temperatures...

Figure 9.9 Two URLs are recommended in reply 1 where user writes about how useful they are.

Thread 19413707

Source: <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/19413707>

Initiating post 2017-02-06 06:46:51 Likes: 2

User n4-1810

I am now really concentrating on getting my 5 a day, my diabetes blood tests this week seem to have stabilized as I am really thinking in advance what I should be eating. I am trying to lean towards lower carbs, but it is difficult. I need to lose 2 stone! how many carbs do you think I should be eating per day? I have my calories down to just below 1000.

Reply 1 (first contribution) 2017-02-06 13:17:16 Likes: 0

User n4-1983

Hi [n4-1810],

If you are seriously thinking of cutting carbohydrate in your diet I would recommend these two websites for guidance. I have found them very helpful and following their advice I have managed to lower my blood glucose and keep it more stable.

www.diabetes.co.uk

www.dietdoctor.com

Good luck

Reply 2 (first contribution) 2017-02-07 10:55:42 Likes: 1

User n4-1810

yes thanks for that I have just purchased the diabetesc.co.uk book on counting carbs and it is extremely helpful and very easy to follow.

In both the initiating post in Figure 9.8 and reply 1 in Figure 9.9, the users explicitly introduce and refer to the URLs posted with “Here is a great piece of research”, “I would recommend these two websites”, similar to the introductory frame used in minimal wording. This discourse indicates the function of information sharing via URLs, and it is indeed interpreted as such by other users, “Thanks very much for the *link!*” and “yes *thanks* for that”.

However, in contrast to the minimal wording, both users also write in length regarding the content of the URLs. In Figure 9.8, after introducing the URL, the user comments on the content linked to the URL, “Apparently, warmed soils and mouldy breads were once used”. In reply 1 in Figure 9.9, the user writes about their personal experience with the URLs, “I have found them very

helpful...I have managed to lower my blood glucose”. This reply also shows that information sharing via URLs can be in response to information seeking by other users (Polletta et al., 2009). This is evidenced by the user’s invitation, “if you are seriously thinking of cutting carbohydrate in your diet”, before introducing the URLs, which is in response to the information seeking in the initiating post “I need to lose 2 stone! How many carbs do you think I should be eating per day?”.

Citing URLs as evidence

In long comments where users write about their stances, the URLs are cited as evidence, similar to the practice of in-text citing in academic writing, which has been shown in the collocation analysis. The example shown in Figure 9.10 illustrates how users include the URLs in their stance-taking, whereas the example in Figure 9.11 illustrates users quoting verbatim from the URLs.

Figure 9.10 The URL is cited as “an example”

Thread 6319879

Source: <https://www.futurelearn.com/courses/soils/1/comments/6319879>

Initiating post 2015-07-18 18:23:59 Likes: 0

User s1-712

Soil security is not just about mitigating the physical effects of nature, it also concerns how we utilise our agricultural products. Animal protein consumption is on the increase and as an example (http://www.earth-policy.org/data_highlights/2011/highlights22) it takes seven pounds of grain to produce one pound of beef. This is obviously an inefficient means of feeding people that will ultimately effect soils. So even though the majority of us don't farm, we do have an impact on how's soils are used.

Reply 1 (first contribution) 2015-07-29 16:00:45 Likes: 0

User s1-30

Not all beef requires this amount of grain, the old hill breeds in the UK lived mainly on grass, hay and straws of various kinds. Welsh blacks for example, and I think also pedigree Herefords.

Reply 2 (first contribution) 2015-07-29 20:17:05 Likes: 0

User s1-419

I have to agree with [s1-419] that the old breeds not fed the same way as intensively farmed cattle/sheep. I also agree with Francis that meat eating is a problem for worldwide food security.

However I put forward the idea that many farms in the UK are very suitable for growing grass and not for crops to directly feed people and feel that animal farming may the best use of these soils.

[5 more replies are omitted]

Figure 9.11 Thread 4423693. The text of the URL is quoted verbatim.

Thread 4423693
Source: <https://www.futurelearn.com/courses/inequalities-in-personal-finance/1/comments/4423693>

Initiating Post 2015-04-08 18:30:43 Likes: 0
User f1-394
Of course there is potential that, with baby boomers now retiring, the political risk is reversed. The older generation having lived their working lives in a low-tax era now have more share of the vote and could increase taxes to maintain their pension payments, at the expense of younger enations.

Although it isn't happening so far, in fact the reverse is true.

Reply 1 (first contribution) 2015-04-09 06:26:55 Likes: 0
User f1-323
I disagree , it is happening . The Coalition have preserved pensions relative to social security & job security for the poor . See Article in The Conversation yesterday “ In his first speech to the Conservative Party Conference as prime minister in 2010, David Cameron drew on the slogan: “We’re all in this together” in the face of the economic challenges facing Britain at that time. Five years later, however, it is clear that it is in fact the poorest who have paid most dearly as a result of various coalition tax and benefit changes.”
http://theconversation.com/state-of-the-nation-inequality-rising-shows-were-not-all-in-this-together-39771?utm_medium=email&utm_campaign=Latest+from+The+Conversation+for+8+April+2015+-+2618&utm_content=Latest+from+The+Conversation+for+8+April+2015+-+2618+CID_7055ba0ee9e742efa0df2f4f84ac5253&utm_source=campaign_monitor_uk&utm_term=State%20of%20the%20nation%20inequality%20rising%20shows%20were%20not%20all%20in%20this%20together

In the initiating post in Figure 9.10, the user expresses their stance, “Soil security is not just about mitigating the physical effects of nature”, and supports their stance with “an example” with an URL enclosed in brackets, then continues their argument, “it takes seven pounds of grains to feed one pound of beef”. This post in turn triggers other users’ stance-taking, as shown in “Not all beef requires this amount of grain” in reply 1, “I have to agree” in reply 2 and other replies not shown in Figure 9.10. This way of including URL in the users’ comment is similar to the in-text citation in academic writing. Rather than using the URL to substitute one’s voice, as has been shown earlier in unaccompanied URLs or minimal wording, the users write in length their own stance. Unlike the

unaccompanied URLs or URLs posted with minimal wording, the URLs cited are not the main content of the longer comments.

In reply 1 in Figure 9.11, the user disagrees with the initiating post, and support their argument that “The Coalition have preserved pensions relative to social security & job security for the poor” with URL by explicitly stating “See Article in The Conversation”, in which *see* and *article* are the collocates of URLs found in section 9.3.1. Although the reply is long, most of the content is verbatim quotation from the URL, and enclosed in quotation marks. There are also times users do not put the verbatim text within quotation marks. Arguably, quoting shows that users distil the most relevant information or evidence from the URLs to support their stance. However, quoting is similar to substituting one’s voice with the URL because it is not the user’s own wording. This contrast with citing URLs as evidence or example, as shown in Figure 9.10, where the URL-posting user writes in length their own opinion and the URL is only used to support their stance. Depending on whether the users also write other content besides quoting, these long comments with quoting can contain mainly the content linked to the URLs.

9.4 Conclusions regarding the extent and general discourse practices of URL-posting

The analysis thus far shows that the URL-posting practices in FutureLearn are similar to those found in other online spaces (Connor, 2013; Polletta et al., 2009; Wikgren, 2001). While users differ in their reliance on URLs, most users are positive to the URLs posted. It is also found that users post URLs to share information, as indicated by their explicit introduction of the URL, whether with little to no elaboration or with longer comments. Besides, URLs are cited in text, listed as a reference or quoted as evidence to support stances. This referencing may in part due to the academic nature of online discussion in MOOCs, while users are also prompted to share URLs by the particular learning design. More importantly, users seem to use URLs to substitute their own writing and represent their voice in the online discussions, further suggesting that hyperlinking is another meaning-making resource

that users can draw upon and make intertextual references to other online spaces. Similarly to on social media, URL-posting is a way for users to express themselves and participate in the wider society (Edgerly et al., 2016; Oeldorf-Hirsch & Sundar, 2015).

Two main findings raise further questions about URL-posting in the FutureLearn online discussions. First, some users rely heavily on URLs in their comments, while the majority never use an URL. This points to the distinct approaches among users when it comes to using sources to support their stances or interact with others in online discussions. There could be a tension between the use of personal opinions and web sources as evidence when users engage in online deliberation, especially when web sources are equated to *hard currency* and authority on one hand and as a made up source compared to authentic experience on the other hand (Polletta et al., 2009; Savolainen, 2014; Wikgren, 2003). Furthermore, different types of internet sources are posted, ranging from those held by established media and professional organizations to social media and user-generated contents. This raises the question of how users evaluate sources posted by others, and how users who employ different types of internet sources engage with each other and co-construct the value of the URLs posted (Connor, 2013; Savolainen, 2014; Wikgren, 2003).

Second, the observation that users differ in their discourse practices of including URLs in their comments suggest that they might interpret or construe URLs differently, and this difference may disrupt their intersubjectivity with each other. As illustrated in the comments containing unaccompanied URLs, minimal wording, or mainly verbatim quote from a URL, URLs can take the center stage of a comment, while other users may interpret the URLs posted differently given that the posting users write so little regarding the relevance of the URLs. All these observations point to the need for micro-analysis that examines how users respond to each other when URLs are involved in their interactions, and how differences in their use of URLs or other information sources underlie the processes of intersubjectivity between them.

In short, in the age of the networked society, online information of various sources can be easily shared and employed by users, while misinformation or partisan ideology can also be easily

propagated (Burnett & Jaeger, 2008; de Maeyer, 2013; Giglietto et al., 2020; Gilster, 1997; Jacobson et al., 2016; Lankshear & Knobel, 2006; Secker, 2017). Therefore, users' conceptualization, usage and evaluation of web sources are an important aspect of digital literacies that warrant further investigation. While the analysis thus far reveals the general positive trend of URL-posting in FutureLearn online discussions, a series of micro-analyses are conducted next to further explore how users co-construct web sources in threads where URLs or the posting of URLs becomes an issue in their stance-taking. This choice of focus also continues the line of inquiry in Chapter 8 that explicates users' discourse practices in processes of intersubjectivity when they disagree. Ultimately, this analysis will address RQ3 as to how URL-posting initiates, sustains or hinders dialogic conversations.

9.5 Micro-analysis

The micro-analysis is organized into three sections, based on the aspects of URL-posting that becomes the center of discussion among users in a thread. First, users debate around the usefulness of the URLs posted, as explored through the example of threads where the most cited source Wikipedia is discussed. Second, URLs are used as the *hard currency*, as termed by Wikgren (2003), for winning a debate when disagreeing users present URLs to challenge or 'overthrow' URLs posted by others. Third, the provision of URLs is construed as necessary when one disagreeing party only employ personal knowledge and experience, and another presents URLs, thus revealing the tension between different types of information sources. In this thesis, users' disagreement on these three aspects is referred to as *link war*.

As revealed in the previous sections, users differ in their reliance on URLs and discourse practices of URL-posting. A link war is waged when this difference comes into conflict. The three aspects of link war are examined by paying attention to how users incorporate the URLs or respond to the URLs posted by others. By investigating link war, the distinctive nature and consequences of URL-posting in online discussions can be better understood, according to the CA principles (Heritage,

2004). The findings reveal that although URLs generate discussion among users, fixation on the URLs may reduce the chance for intersubjectivity as users focus not on negotiating the content being shared but on the authority and ideological positioning of the source itself.

9.5.1 Co-constructing the value of Wikipedia as a source

This section explores users' co-construction of the value of Wikipedia, the most cited source in the online discussion in FutureLearn, as revealed in section 9.2.3. Most of the time, users appreciate sharing of Wikipedia, as shown in Figure 9.12. However, there are a few times that Wikipedia is subjected to a discussion among users, as shown in Figure 9.13. In these instances, through their interactions, users co-construct the value of Wikipedia, showing that the value of a source can change depending on the communicative context.

9.5.1.1 "Thank GOD for *Wiki!*"

In the thread in Figure 9.12, a user posts a Wikipedia URL to clarify a concept taught in the course video, which receives replies from five users, all of whom thank the user for posting this URL, which is deemed as superior to the course video.

Figure 9.12 All the replying users thank the URL-posting user for sharing the Wikipedia link.

Thread 529094

Source: <https://www.futurelearn.com/courses/moons/1/comments/529094>

Initiating post 2014-03-19 12:00:58 Likes: 10

User m1-185

Don't confuse the model with reality. As well as eccentric orbits, the three moons of Jupiter can never all be in conjunction (aligned) at the same time, only two at a time. According to http://en.wikipedia.org/wiki/Orbital_resonance (which answers most questions posted here), this is a remarkable relationship; it isn't true of all triple resonances.

Reply 1 First reply 2014-03-19 15:22:24 Likes: 0

User m1-1090

Thanks for the link [m1-185] the model did look too neat to be absolutely true but it was (as they say) fit for purpose. Loved the link you suggested (Thank GOD for Wiki !) for me it complemented and ameliorated the OU text and visual. [m1-1090]

Reply 2 First reply 2014-03-20 12:44:38 Likes: 0

User m1-491

Thanks [m1-185] for the link. Expanded on the video.

Reply 3 First reply 2014-03-20 15:58:52 Likes: 1

User m1-665

Thanks for the link [m1-185], I don't think the above video does do the job, it is misleading and arguably outright wrong as without the clearer wiki bit I would have argued the point that they did line up on the basis this course is factual, and such facts should be correct. Adding the Jupiter and its moon names means the animation above should reflect reality.

Reply 4 First reply 2014-03-21 17:53:42 Likes: 1

User m1-1536

I agree with [m1-665]. Someone goofed up bigtime in creating this misleading and incorrect demonstration. Thanks for setting us straight, [m1-185].

Reply 5 First reply 2014-03-22 23:00:07 Likes: 0

User m1-1452

Really helpful thank you [m1-185]!

Reply 6 Initiator's reply 2014-03-25 11:43:43 Likes: 2

User m1-185

I don't think anyone goofed up here. They simplified the model to make the resonance clearer to the unlearned. Job done.

Let me generalise: NEVER confuse ANY model with reality. Too many people do, especially the press and politicians when it suits them. And, unfortunately, central banks. :(By definition, a model is a simplification.

In the initiating post in Figure 9.12, user m1-185 takes a stance “Don't confuse the model with reality” and supports this claim with a Wikipedia URL following “According to”. All the other users who contribute in this thread thank the URL-posting user, as shown in reply 1 to 5, where *thank(s)* is mentioned in all replies. The value of the Wikipedia URL posted is emphasised in reply 1 “Thank GOD for Wiki!”. The value of Wikipedia is collaboratively co-constructed by users in this thread, as each reply seems to build on the previous one. In reply 1 and 2, the Wikipedia URL is construed as an addition to the course material, “it complemented and ameliorated the OU text and visual”, “Expanded on the video”. This positive sentiment on Wikipedia is further built on by reply 3 and 4 that contrast it to the course content which “is misleading and arguably outright wrong”, “misleading and incorrect”. The users contributing reply 1 to 4 co-construct the value of Wikipedia by not only expressing appreciation towards the URL-posting user, but also contrasting the site to course materials which are being painted negatively. Nonetheless, the URL-posting user neutralizes this contrast in reply 6 by saying the course video “simplified the model to make the resonance clearer to the unlearned. Job done.” Users’ positive co-construction of Wikipedia, as illustrated in this thread, may explain why Wikipedia is most used in the online discussions.

9.5.1.2 “You should be wary of using *Wikipedia*”

However, Wikipedia is not always construed as positive. In the thread in Figure 9.13, the user who posts the Wikipedia URL is cautioned by another user for using Wikipedia, yet both agree the function of Wikipedia as an “initial” or “secondary” source. This thread is one of the four threads found in the corpus in which users engage in meta-discussions around Wikipedia and construe Wikipedia as a good start for information but not a reliable source.

Figure 9.13 Discussion about Wikipedia

Thread 643808

Sources: <https://www.futurelearn.com/courses/moons/1/comments/643808>

Discussion prompt: What measurements would you most like it to make when embedded in the ice of a moon such as Europa?

Initiating post 2014-04-10 19:35:35 Likes: 4

User m1-1030

Surely, if the 'mother' craft was in orbit the shell would need to decelerate rather than accelerate! (Or, for the purists: Accelerate in the opposite direction to the 'mother craft', relative to the moon.)

[two paragraphs are omitted]

It's not rocket science is it? Oh...

Reply 1 First reply 2014-04-17 13:57:16 Likes: 0

User m1-1003

The (surface) escape velocity of Europa is about 2.0 km/s. The speed at which a shell would impact if simply dropped in a suitable Jupiter orbit is therefore _at least_ 2 km/s plus whatever relative speed you fire it at. Given that it is apparently designed for a 300 m/s impact, the shell would not survive. You'd have to fire the impactor _away_ from the moon.

[one paragraph is omitted]

(Insert equally snide remark here.)

Reply 2 First reply 2014-04-24 09:55:37 Likes: 2

User m1-1048

Surely you've got something wrong here [m1-1003]. The figures you give make no reference to altitude. Therefore, according to your analysis, if the shell were released at an altitude of 1mm, it would still hit the surface at about 2 km/s.

[one paragraph is omitted]

Reply 3 Further reply 2014-04-28 14:27:19 Likes: 0

User m1-1003

http://en.wikipedia.org/wiki/Escape_velocity

Reply 4 Further reply 2014-04-29 12:29:25 Likes: 1

User m1-1048

[m1-1003] - I'm not querying what escape velocity is, but what you're trying to use it for. The clue is in the name, 'Escape' - the velocity to be reached to enable a body to continue to move away without the aid of a propulsion unit. The escape velocity has nothing to do with the impact velocity if something is dropped - apart from the fact they will both be influenced by factors such as mass, gravitational constant and distance.

[one paragraph is omitted]

Reply 5 Further reply 2014-04-30 07:30:37 Likes: 0

User m1-1048

Incidentally [m1-1003] you should be very wary of using Wikipedia as a reference source. Its open nature means that entries can be altered by virtually any one. In the past it has been altered by students as a prank, "cranks" who think the scientific world has got it all wrong and for political reasons - this morning's news carried an article about a Wikipedia page on Muslims being altered.

If you are going to use Wikipedia as your initial source you should also check it with a more reliable source. In the case of escape velocities I don't think there is a problem as the Wikipedia page gives the same values as the NASA site (but I only checked the entries for Earth and Europa).

[Reply 6 contributed by another user is omitted because it is not relevant to the current analysis]

Reply 7 Further reply 2014-05-02 19:20:57 Likes: 0

User m1-1003

[m1-1048], the link is simply a reference to a place where your question is answered. I do not rely on Wikipedia as an academic source so never fear. (Vandalism also rarely affects such types of article, for what it's worth.)

Since you asked nicely, here are some other references for both the concept and the number:

J.R. Wertz, Orbit & constellation design & management. 2001. Microcosm, Hawthorne, CA, USA & Springer, New York, NY, USA. pp. 45, 845-847, 854.

G.P. Sutton & O. Biblarz, Rocket propulsion elements. 8th ed. 2010. Wiley, Hoboken, NJ, USA. pp. 116-118.

J.R. Wertz, Orbits and astrodynamics. 2011. (In J.R. Wertz, D.F. Everett & J.J. Puschell, eds, Space mission engineering. Microcosm, Hawthorne, CA, USA). p. 201, 205.

N. Sarzi-Amade, Physical and orbit properties of the sun, earth, moon and planets. 2011. (In J.R. Wertz, D.F. Everett & J.J. Puschell, eds, Space mission engineering. Microcosm, Hawthorne, CA, USA). p. 955.

(Note that table 4-1 on page 119 of Sutton & Biblarz is way out of date, so don't use those escape velocities for anything. The authors are aware and it should hopefully be fixed in the 9th ed.)

In Figure 9.13, user m1-1003 posts an unaccompanied Wikipedia URL in reply 3. This URL display is in response to reply 2 in which another user m1-1048 considers m1-1003 as having “surely ... got something wrong”. This unaccompanied URL may serve as substituting user m1-1003’s voice, as explained by the user themselves in reply 7, “where your question is answered”. This unaccompanied URL is similarly interpreted by user m1-1048 as a response, as evidenced in reply 4 “I’m not querying what escape velocity is” which is what the URL is about, as indicated by the URL address. Consistent with previous analysis, the unaccompanied URL in reply 3 seems to carry the function of representing one’s voice or as evidence to counter the other’s stance.

More importantly, the interactions between user m1-1003 and m1-1048 in reply 3, 5, 7 reveal their co-construction of the Wikipedia as secondary to “reliable source” and “academic source”. In reply 5, m1-1048 construes Wikipedia as “*initial source*” that users “*should be very wary*” and “*check it with a more reliable source*”. In response, in reply 7, m1-1003 construes Wikipedia “as simply a reference” and not “an academic source”. The mention of “NASA site” by user m1-1048 in reply 5 and the listing of references by m1-1003 in reply 7 also contrast with Wikipedia which appears less “reliable” than these “reliable” or “academic” sources. The listing of references in reply 7 provided by the same user who first posts the Wikipedia link also suggests that the user might only use the Wikipedia URL as a quick and easy way to substitute their own words to respond. However, when it fails, they thus make the effort to provide “academic source”, although not necessarily elaborate on any of the references, as can be seen in reply 7. Interestingly, although there is no request from user m1-1048, user m1-1003 introduces the list of references with “Since you asked nicely...”, possibly using the reference lists as evidence for one’s “knowing” status rather than for discussions, after being cautioned for using Wikipedia. This indicate that evidence is sometimes presented to establish one’s epistemic status instead of claim. Lastly, the contrast between Wikipedia and academic sources made by these users reveals the tension between popular sources and scientific sources (Savolainen, 2014), and users seem to employ them for different functions in

the online discussions, and for different communicative purposes. In this case, Wikipedia URL as a quick answer, academic sources to legitimate one's epistemic status.

Contrasting responses toward Wikipedia URLs posted in Figure 9.12 and Figure 9.13 attest to individual differences in preferences for information sources, and possibly the value of the Wikipedia as a source may vary depending on the topic at stake. Nonetheless, both threads show that Wikipedia may be a source for simplicity, consistent with Singer et al's (2017) findings that Wikipedia is generally relied on by internet users for an overview of a subject. Users' discussion about Wikipedia also mirror other situations where users mention preference for academic sources over URLs posted by other users, "thank you for the link you've just shared but if you want to exchange real science papers..."¹⁵⁸, or queries for credential, "I'd be wary of those links you posted particularly Jillian Michaels - qualifications?..."¹⁵⁹. This suggest that users assess the URLs posted and sometimes engage in meta-discussions around the URLs, and co-construct the value of the URLs for the particular topic they are discussing. They attribute different authoritative value to different sources, and the value may change depending on their purposes (Connor, 2013; Flanagin & Metzger, 2000; Singer et al., 2017; Wikgren, 2001). The meta-discussion about the URLs posted, as shown in the discussion of Wikipedia, despite off-topic, show that users co-construct their idea on information sources in their interactions with each other. However, the co-construction of the value of the URLs posted may not always occur, as shown in the next two sections when each user sticks to the URL posted by themselves, and explicitly critiques the URLs posted by others.

9.5.2 Using URLs to respond to URLs

This section explores the tension when users post URLs linked to different online sources within their interactions, thus extending the quantitative findings in section 9.2.3 that found various types of online sources have been posted in the online discussions. User-user interactions within a thread can

¹⁵⁸ <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/19005625>

¹⁵⁹ <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/18950941>

sometimes evolve into URL-URL interactions when users with opposing stances each presents URLs to support their stance and overthrow the others' posted URLs. The credibility of the URLs posted by each user becomes the point of contention and the weighting for an argument, instead of the information contained in the URLs. Rather than being cited as evidence for claim, the URL itself becomes the evidence, in other words, hard currency as observed by Savolainen (2014) and Wikgren (2003). This is illustrated in two examples below, where users' responses to URLs are examined, when different sources are posted within the same threads, and are used to support opposing stances.

9.5.2.1 "your one posted *link*" vs "My *link* carefully explains"

This thread is a long thread with 26 replies initiated by user ah1-639 and is on human faeces as fertiliser, then evolves into the effect of glyphosate, which user ah1-639 and ah1-993 disagree on. The disagreement between the two users evolves from the topic glyphosate to a URL-URL interaction and a discussion on the credential of experts mentioned in the URLs posted by both users. The URL-URL interaction starts when ah1-993 requests ah1-639 for URLs. Each of them comes back with URLs to argue against the URLs posted by the other. Their link war happens between reply 10 to 20 of this thread as shown in Figure 9.14 and Figure 9.15.

Figure 9.14 First part of thread 20311486

Thread 20311486

Source: <https://www.futurelearn.com/courses/ancient-health/1/comments/20311486>

[The initiating post and reply 1 to reply 9 are omitted]

Reply 10 (Initiator's subsequent contribution) 2017-03-12 16:25:42 Likes: 2

User ah1-639

And the argument becomes more complex in that we need to ask about the status of the feed that the animals ate - was it sprayed with glyphosate-based products? If so, that herbicide will travel through the animal to our plates. Already studies are showing a wide range of autoimmune conditions linked to this practice. Sadly, most farmers have no idea about the status of their feed as most no longer grow their own but instead purchase it at the local farm store. Moreover, glyphosate is only one of many chemicals that have been shown to have profound effects upon humans, animals, aquatic life, etc. This of course leads to wider concerns such as food security, seed variety loss, soil nutrient depletion, the list goes on.

[Reply 11 is omitted]

Reply 12 (subsequent contribution) 2017-03-12 17:26:38 Likes: 0

User ah1-993

I wouldn't have a problem with glyphosate as it is a much safer herbicide than many we have used over the last few decades. I'm not sure that there is any *good* evidence that its use is linked to autoimmune disease (*happy to be proved wrong if you can link to the studies*).

Reply 13 (Initiator's subsequent contribution) 2017-03-13 16:36:51 Likes: 0

User ah1-639

Here's a start; plenty more out there if you search:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3945755/>

<https://search.informit.com.au/documentSummary;dn=641406900036750;res=IELNZC>

<http://www.neuroregulation.org/article/view/15833>

<http://www.hoajonline.com/autism/2054-992X/3/1>

<http://notoxicliving.com/wp-content/uploads/2015/11/Glyphosate-pathways-modern-diseases.pdf>

http://farmwars.info/wp-content/uploads/2016/07/Glyphosate_pathways-to-modern-diseases-V-Amino-acid-analogue-of-glycine-in-diverse-proteins_FNL_Published.pdf

Reply 14 (subsequent contribution) 2017-03-13 17:17:29 Likes: 1

User ah1-993

It isn't so much about searching and linking every mention of glyphosate [ah1-639], but about gathering credible evidence.

Three of your links are to articles by Seneff, a notorious anti-GM activist. She has a degree in electrical engineering. Her co-author, Samsel is a research scientist interested in pollution. They have no biological or medical expertise, they did no research themselves - it's all speculation.

You might be interested in the response from medics

<https://www.biofortified.org/2015/01/medical-doctors-weigh-in-on-glyphosate-claims/>

The 3rd link claims that glyphosate use has risen over the last 25 years and so has autism, diabetes and coeliac disease, so we should stop using glyphosate. It doesn't show that glyphosate causes these conditions. After all, consuming organic food has also risen over the last 25 years, so if this was a valid argument we should also stop eating organic food because it correlates with an increase in autism, diabetes and coeliac disease.

Your 2nd, 4th and 5th links are about polycystic ovary syndrome (again a hypothesis not a cause), autism and cancer - these are not autoimmune diseases.

Reply 15 (Initiator's subsequent contribution) 2017-03-14 00:10:57 Likes: 0

User ah1-639

Thanks for your response, [ah1-993]. However, *your one posted link is from two doctors who spend a large portion of their time debunking anything that does not agree with their world view.* I am not interested in an argument so I will *agree to disagree* with your opinions on this subject. I hope that you will too.

Note. User ah1-639 posts a list of URLs in reply 13 after being requested by user ah1-993 who in turn posts a URL in reply 14 in response to the list of URLs.

Figure 9.15 Second part of thread 20311486

[Reply 16 is by another user which is not relevant to the current analysis, so is omitted]

Reply 17 (subsequent contribution) 2017-03-14 07:25:38 Likes: 2

User ah1-993

Hi [ah1-639], I'm not looking for an argument either, just *accurate information*, but *debunking anything that does not agree with her world view is exactly what Seneff does!* My link carefully explains why Seneff has misrepresented the work of other scientists. It does quote the two doctors you mention, but *is actually written by Karl Haro von Mogel, a plant geneticist, for Biology Fortified, a website that aims to inform the public about biotechnology and other issues in food and agriculture through science-based resources.*

Like most people, I would love to see a world where we did not have to use pesticides and herbicides, but I appreciate that they are necessary in large scale farming. One of the reasons why glyphosate is popular is because it is one of the safest herbicides we have.

[One paragraph in this reply is omitted, because it is in response to reply 16 which is omitted as well, but not relevant to current analysis]

Reply 18 (Initiator's subsequent contribution) 2017-03-16 14:49:25 Likes: 0

User ah1-639

I remain unconvinced. First, the bulk of the research done on herbicides and pesticides is industry funded which means that commercial interests are at stake. Second, in viewing the writings of the authors you cite above (Dr Steven Novella & Dr David Gorski aka Orac) I note that they spend an inordinate amount of time defending Monsanto. Why, I wonder? Third, new information re: the safety of glyphosate is out, see:

https://www.nytimes.com/2017/03/14/business/monsanto-roundup-safety-lawsuit.html?emc=edit_nn_20170315&nl=morning-briefing&nid=79137994&te=1&r=0

Like many 'safe' and 'useful' products before it (mercury, thalidomide, DDT, asbestos, to name but a few), the use of Glyphosate will no doubt come under increasing independent challenge and will hopefully be replaced by something less problematic. Changing farming practice would be a good start but that's an entirely different topic.

Skeptics are easy to find, locating those who will stand up to corporate greed is sadly, a much more difficult task.

Reply 19 (subsequent contribution) 2017-03-16 17:47:06 Likes: 0

User ah1-993

Coincidentally, I have spent some time this afternoon spraying glyphosate - the first time in my life I have used it, but my elderly mother wanted the moss and weeds on her drive killed as they were a dangerous slip hazard for her.

I don't use pesticides and herbicides in my own garden, but I'm comfortable using glyphosate because I know about the many independently funded studies that have failed to show adverse health effects. And I trust scientists and researchers who are skilled at evaluating the evidence more than I trust journalists with an agenda.

As practicing medical doctors it would be odd if Novella and Gorski had a particular reason to defend Monsanto (can you give examples?) other than calling out pseudoscience when it occurs, especially since glyphosate is now off patent and produced by many manufacturers and companies worldwide.

Reply 20 (subsequent contribution) 2017-03-16 17:49:49 Likes: 0

User ah1-993

Dr Novella has discussed the NY Times article.

He concludes

'I am simply searching through PubMed to find reviews of the safety of glyphosate, and this is what I find. You can do the same, it's a user-friendly searchable database. There is a remarkable consistency to the reviews - they all agree that the evidence does not support an association between glyphosate exposure and any adverse health outcome... Glyphosate, in fact, is one of the safer pesticides in use (including many organic pesticides). It has replaced far more toxic herbicides. Opposing glyphosate because of unwarranted fears of toxicity is likely to cause harm due to whatever replaces it. Tilling is bad for the soil and releases CO₂ into the atmosphere, and we cannot feed the world through hand weeding. Herbicides have to be part of the equation, and glyphosate is one of the safest out there.'

<http://theness.com/neurologicablog/index.php/does-glyphosate-cause-cancer/>

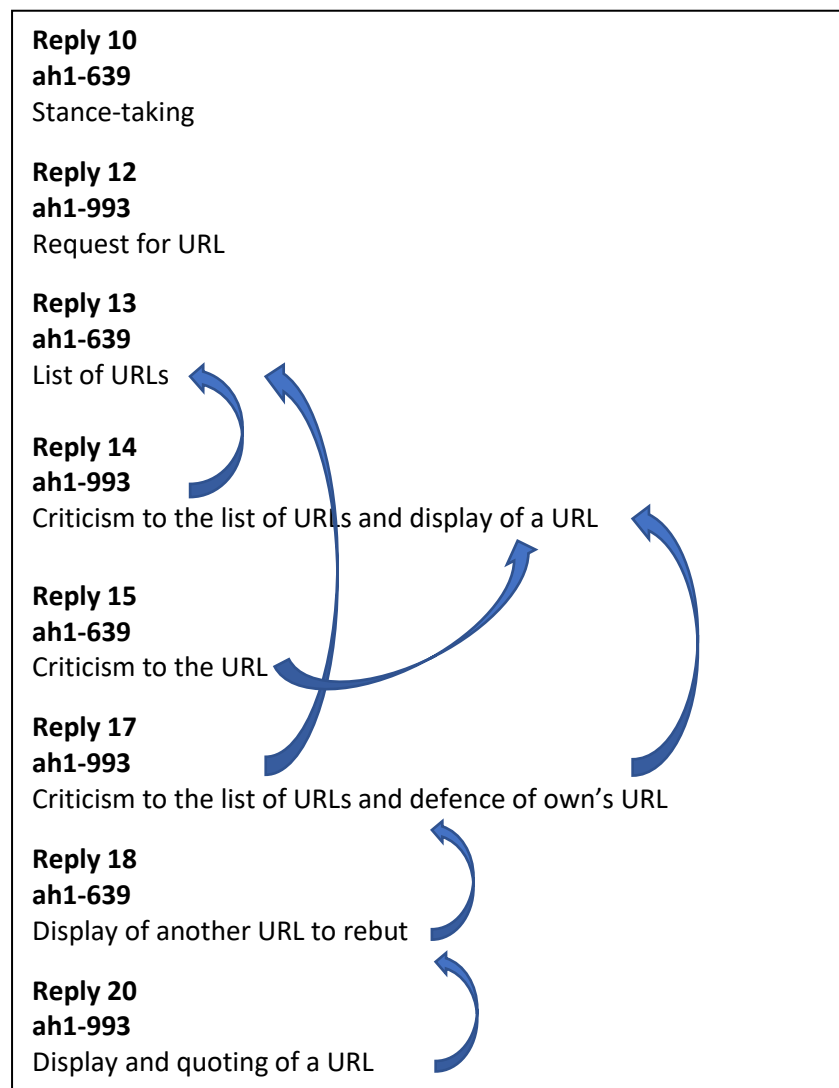
[6 more replies are omitted]

Note. User ah1- 993 posts an URL in reply 20 in response to ah1-639's URL in reply 18.

In this thread, two URL-URL interactions can be identified where an URL is used to respond to an URL posted by another between reply 13 and 14 (Figure 9.14) and between reply 18 and 20 (Figure 9.15). In reply 13, ah1-639 lists six URLs upon request by ah1-993 in reply 12 “happy to be proved wrong if you can link to the studies”. The links are critiqued by ah1-993 one by one in reply 14, “Three of your links are to articles by Seneff, a notorious anti-GM activist.”, “The 3rd link doesn’t show that glyphosate causes these conditions”, “Your 2nd, 4th and 5th links..... are not” about “autoimmune diseases”. More importantly, ah1-993 also posts a URL saying, “You might be interested in the response from medics”. The URL contains criticism against the authors of some of the articles linked to the URLs posted by ah1-639. Therefore, the URL posted by ah1-993 can be interpreted as a direct rebuttal to the URLs posted by ah1-639. In fact, in reply 14, ah1-933’s critique is mostly based on this URL.

The second URL-URL exchange happens when ah1-639 remains “unconvinced” in reply 18 and lists three points to rebut what has been mentioned by ah1-993. The third point introduces an URL with minimal wording “Third, new information re: the safety of glyphosate is out”, indicating it is an evidence to counter-argue. Interestingly, ah1-993 posts an URL that “has discussed the NY Times article” in reply 20, another evidence that a URL is used to rebut another URL. Therefore, the URL posted by ah1-639 in reply 18 can be said to be in response to ah1-993’s reply 14 and 17, whereas the URL posted by ah1-993 in reply 20 can be said is in response to ah1-639’s reply 18. To some extent, it is surprising that users manage to find URLs which seem to directly address the URLs posted by others. This mirror the fact that the information on the web can be contradictory. The URL-URL interaction is visualized in Figure 9.16.

Figure 9.16 The characteristic of the URL-URL interaction in thread 20311486.



Note. The arrows indicate the relationship between the URLs posted by the two users.

The link war is not limited to using URLs to respond to another's URL, but also users' criticism towards each other's URLs in terms of the ethos of the authors or experts mentioned in the URLs, instead of the content. In reply 14, ah1-993 critiques the author of the three URLs posted by ah1-639 as "a notorious anti-GM activist" who "has a degree in electrical engineering" and "no biological or medical expertise", and "did no research themselves". This critique suggests that users may attribute expertise knowledge to some URLs but not all URLs. Similarly, in reply 15, ah1-639 critiques the experts mentioned in the URL posted by ah1-993, "two doctors who spend a large

portion of their time debunking anything that does not agree with their world view". Similar criticism is repeated in reply 17 and 18. In reply 17, ah1-993 applies what ah1-639 writes in reply 16 onto the authors of the URLs posted by ah1-639, "debunking anything that does not agree with her world view is exactly what Seneff does!" In reply 18, ah1-639 also queries the experts mentioned in the URLs posted by ah1-993, "they spend an inordinate amount of time defending Monsanto". The criticism towards the authors or experts mentioned in the URLs posted suggest that users seem to pay more attention to the authority of the URL itself rather than directly engaging with the information and content. This focus is probably because both have presented URLs as evidence for their claim, such that the 'battleground' is elevated to evaluating the credibility of the URLs posted by each other.

Besides critiquing others' URLs, ah1-993 also defends their own URLs by highlighting the credential of the author and aim of the website, as shown in reply 17 "written by, a plant geneticist, for ... a website that aims to inform the public about biotechnology and other issues in food and agriculture through science-based resources." The user also defends the experts mentioned in the URL, as shown in reply 19 "As practicing medical doctors it would be odd if Novella and Gorski had a particular reason to defend Monsanto". The user's defence of their own link is also evidenced in the positive evaluation, "My link carefully explains". The criticism and defence of the authors or experts mentioned in the URLs shows that the authority of an URL as evidence rests on the credibility and credentials of its authors or experts mentioned in the URL. This can be an indication of appealing to authority (Savolainen, 2014). The reliance on URL as evidence is also highlighted by ah1-993's request of URLs in reply 11 "happy to be proved wrong if you can link to the studies", suggesting that one will only change their stance if there is a link. This user further proposes that URLs are about "credible evidence", "accurate information", and "scientists and researchers who are skilled at evaluating the evidence", rather than "journalists with agenda". This might be similar to users in diabetes newsgroup who use abstracts of medical journals, which are

considered as scientific sources to legitimate their claims, and criticise other sources (Wikgren, 2003).

The discourse practices of how users incorporate URLs in their comments also indicate how they construe the URL itself as evidence. In the conversations between user ah1-993 and ah1-639, instead of in-text citation, every URL is introduced explicitly with minimal wording, despite being posted in long comments. For example, “You might be interested in the response from medics” in reply 14, “new information re: the safety of glyphosate is out” in reply 18. In reply 20 which mainly contains only quotation from an URL, the only wording by the user is highlighting the expert being quoted, “Dr Novella has discussed the NY Times article”. In contrast, no citation is provided for “the many independently funded studies that have failed to show adverse health effects” in reply 19, although this user, ah1-993, requests “link to studies” from ah1-639, and posts 92 URLs in the discussion of this MOOC. The discourse practice of introducing URLs as authoritative evidence suggests that URL itself becomes an evidence, and equivalent to expert knowledge, rather than its content being drawn on for elaborating on stance in the case of in-text citation. This is similar to what Wikgren (2003) finds in a diabetes newsgroup where users seldom express the need of reading into the details of the sources, suggesting that referring to the web sources may only be a rhetorical strategy to win an argument.

The disagreement between the two users seems to be irreconcilable, as indicated by user ah1-639’s “agree to disagree” mentioned in reply 15 (and in reply 24 which is not shown here). This is possibly because both stand by the URLs they posted, as evident by the link war. This points to the possible problem with over-reliance on URLs as evidence, especially when disagreeing parties pay attention to their own URLs and the experts mentioned in the URLs instead of the relevance of the content of the URLs, such that the URLs themselves become the centre of discussions. Given that scientific evidence or expert knowledge can be in conflict at times, so it is not surprising that disagreeing users are able to present URLs supporting opposing views. Given users construe URLs as representing expert knowledge, it might be hard for users posting URLs to be open to the URLs with different

views, thus a link war might not be easily resolved. They might remain polarized, given that the focus on the presence of URLs itself mirrors the situation of “he said, she said”. This might partly explain why users in the Facebook group examined by Jacobson et al. (2016) only post links supporting their stances. The presence of URLs that support one’s stance may also explain why some people stick to their stance despite other evidence being available.

9.5.2.2 Repeatedly posting the same URL

This thread takes place in the discussion space of the nutrition-4 course. The URL-posting practices in this thread mirror other health forums where internet resources are commonly used and exchanged (Wikgren, 2001). However, the exchange of URLs in this thread becomes a link war when users critique others’ URLs and stick to their own URLs. Three users who post URLs in this thread hold different views towards fat and health, two hold strong views for and against and one is on middle ground. Significantly, the user who is against fat in diet posts the same URLs three times in the thread. As in the previous example, the discussion evolves from the topic to criticism towards the URLs posted.

Figure 9.17 First part of thread 18980719

Thread 18980719

Source: <https://www.futurelearn.com/courses/nutrition-wellbeing/4/comments/18980719>

Initiating post 2017-01-19 17:27:45 Likes: 0

User n4-2511

I only use coconut oil to cook with. I take a tablespoon daily (when I remember). I avoid margarine and butters. [.....]

Reply 1 (first contribution) 2017-01-19 19:33:50 Likes: 1

User n4-3342

I've starting using coconut oil also. But reading it, i found out that it have a lot more saturated fat then oil from corn, sunflower and canola! One spoon have around 13g of saturated fat, which correspond to almost 60% of our daily recommendation! =o

Reply 2 (first contribution) 2017-01-19 21:04:23 Likes: 0

User n4-211

Eat it. Much better for you than seed oils and polyunsaturates.. <https://www.ncbi.nlm.nih.gov/pubmed/24723079>

Reply 3 (first contribution) 2017-01-20 19:57:09 Likes: 0

User n4-1657

Doctors have been reverting (yes REVERTING) diabetes-2 and heart disease through avoiding all high fat sources. See e.g. fig 2 in http://dresselstyn.com/JFP_06307_Article1.pdf See <http://drmcDougall.com/> or e.g. <http://pcrm.org/> From this it would appear we are really a low fat species.

Reply 4 (first contribution) 2017-01-23 14:42:45 Likes: 0

User n4-2611

[n4-211], if you're referring to coconut oil, the British Nutrition Association, in October 2016 recommended the low consumption of coconut oil as it's high in saturated fats and so far there's no evidence of its health benefits <https://www.nutrition.org.uk/attachments/article/998/Coconut%20oil%20FAQ%20branded.pdf> Like with everything, I think moderation is best.

Reply 5 (subsequent contribution) 2017-01-24 09:09:43 Likes: 1

User n4-211

There is no problem with satirated fats. This is part of the problem. There is no conclusive evidence saturated fat is harmful. This is currently being challenged to PHE and other nutritional advisories as their recommendations are not backed by research. Heart health is not affected by saturated fat either. Most of this comes from poor science initially done in 1970s by Ansel Keyes. see <http://articles.mercola.com/sites/articles/archive/2016/06/05/saturated-fat-heart-disease-risk.aspx>

Reply 6 (initiator's subsequent contribution) 2017-01-24 12:49:33 Likes: 0

User n4-2511

Many health care advisors advocate eating virgin/organic coconut oil for it's health benefits even though it is high in saturated fats. Coconut oil contains lauric acid, which is a medium-chain fatty acid, that converts to monolaurin. Monolaurin is the compound found in breast milk that strengthens a baby's immunity, and a great deal of research has been done to establish the ability of lauric acid to enhance immunity. This medium-chain fatty acid (MCFA) actually disrupts the lipid membranes of offending organisms such as yeast, fungal and bacteria living in our gut. This is the main reason why I consume it.

Note. URLs are posted in reply 2, 3, 4 where the users hold different views on fat and health.

Figure 9.18 Second part of thread 18980719

Reply 7 (subsequent contribution) 2017-01-25 11:33:10 Likes: 0

User n4-2611

I know about the poor evidence in saturated fat but as I understand there's not enough research (as far as I know) done on coconut oil to prove its benefits.

Reply 8 (subsequent contribution) 2017-01-25 23:04:30 Likes: 1

User n4-2611

Also, I have found a very interesting article discussing the systematic reviews carried out on research in this field which I think it's worth looking at. <http://www.cebm.net/evidence-really-not-support-introduction-low-fat-dietary-guidance-1983/>

It may well be the case that saturated fat is not that bad for you but for now I'll take the WHO, Public Health England and the British Heart Foundation's advice :-)

Reply 9 (subsequent contribution) 2017-01-28 11:51:38 Likes: 0

User n4-1657

"Like with everything, I think moderation is best."

Please realize that we will only know the true meaning of moderation if we know the extremes: the healthiest food and the unhealthiest food. Without knowing the extremes "moderation" means putting your head in the sand.

"There is no problem with saturated fats."

Sorry, if we can REVERT both heart disease and diabetes-2 through a truly low fat diet than there is a definite problem with fats, including saturated fats.

Reply 10 (subsequent contribution) 2017-01-28 11:55:47 Likes: 0

User n4-1657

"There is no conclusive evidence saturated fat is harmful" Are you sure??? Where do you get that from??? If you can revert heart disease AND diabetes-2 through a truly low fat diet (including very low saturated fats) than I would think that strongly indicates that fats are not just bad but really bad. I wonder if the critics of Ansel Keys really understand a truly low fat diet.

Reply 11 (subsequent contribution) 2017-01-28 12:07:48 Likes: 0

User n4-1657

Here is a good presentation on fat and health research: <https://www.youtube.com/watch?v=LbtwwZP4Yfs>

Reply 12 (subsequent contribution) 2017-01-29 08:24:30 Likes: 0

User n4-211

Some of the quoted studies here are from 2000 and old research. There was the presence of high carb too which have more recently been indicated to cause very low density lipoprotein which are the "bad" part of ldl. Also these were funded by big pharma in view of supporting statin sales. Sorry not convinced.

Reply 13 (subsequent contribution) 2017-01-30 20:21:16 Likes: 0

User n4-1657

The newer the research, the bigger the commercial influence. If "Doctors have been reverting (yes REVERTING) diabetes-2 and heart disease through avoiding all high fat sources. See e.g. fig 2 in http://dresselstyn.com/JFP_06307_Article1.pdf See <http://drmcDougall.com/> or e.g. <http://pcrm.org/> From this it would appear we are really a low fat species." does not convince you than nothing will.

Note. User n4-1657 reposts the URLs posted before.

The link war in this thread occurs when three users holding different views towards “fat” post URLs as evidence to support their stance in reply 2, 3 and 4. In reply 2, user n4-211 leaves a URL after the claim “Eat it. Much better for you than seed oils and polyunsaturates [URL]”. This URL is like an unaccompanied URL as there is no elaboration about it, although it could be argued that it is used in a way like in-text citation to support the user’s claim. This lack of elaboration attracts another user, n4-2611, to query if the user is “referring to coconut oil” in reply 4, suggesting that the URL posted is subjected to different interpretation. In reply 4, user n4-2611 also posts a URL after stating the conclusion from the article linked to the URL, “British Nutrition Association [.....] recommended the low consumption of coconut oil as it's high in saturated fats and so far there's no evidence of its health benefits [URL]”. This is similar to in-text citation and with clearer elaboration compared to n4-211’s URL-posting. Another user n4-1657 also seems to respond to reply 2’s claim of “Eat it” with an opposing view, claiming “avoiding all high fat sources” with three URLs posted and strong stance, as shown in the repetition and capitalization, “Doctors have been reverting (yes REVERTING) diabetes-2 and heart disease” in reply 3. These three users all present URLs to support their own stance while disaligning with each other, suggesting that URLs is one way of evidencing one’s stance.

After the initial stance-taking of these three users, user n4-211 comes back in reply 5 with another URL to support their stance, “There is no problem with satirated fats”. The user elaborates on the URL before introducing it with “see....”. This is supported by user n4-1657 in reply 8 “It may well be the case that saturated fat is not that bad for you”, which is a conclusion made after the user shares another URL, “a very interesting article discussing the systematic reviews”. However, the exchange between these two users between reply 5 and reply 8 attracts user n4-1657’s strong objection, as shown in their response in reply 9 and 10, “Please realize that”, “Are you sure??? Where do you get that from???” “I wonder if the critics of Ansel Keys really understand a truly low fat diet”. After their objection in reply 9 and 10, the user further posts a URL in reply 11 with a positive evaluative introductory frame, “Here is a good presentation on fat and health research”.

Posting an URL at the end of one's objection to others' stance again suggests the use of URL as evidence. The frequent URL exchange among these three users in their stance-taking create a link war among them.

Interestingly, within the thread, the initiator also responds to the discussions in reply 6 by writing "the main reason why" they "consume it", but without posting any URL. This reply without any URL is not referred to at all in the thread, possibly being ignored by these three users who have been posting URLs. This raises the concern that focus on URLs can overshadow replies containing no URL in a thread.

Towards the end of this thread, the discussion also evolves into a criticism and defence of URLs, as shown in reply 12, "Some of the quoted studies here are from 2000 and old research", "Also these were funded by big pharma in view of supporting statin sales" and reply 13 rebuttal, "The newer the research, the bigger the commercial influence". These two replies are also where the link war comes to a stalemate when each stick to their stance. This stalemate is evinced in reply 12 "Sorry not convinced.", and reply 13 "If does not convince you than nothing will". Interestingly, in reply 13, user n4-1657 reiterates the same stance and URLs in the same wordings, suggesting they are sticking to their own stance and URL. This user in fact also posts similar claims and URLs in another independent post and reply to another post in the same step.

As with the previous thread, this thread illustrates a user-user interaction which largely builds on a URL-URL interaction, such that reply without URL is not taken up and the thread moves towards criticism and defense of URLs. The obvious stalemate at the end of the thread further point to the problem of reliance on URLs and link war. Each user seems to take URLs as the hard currency for their stance, such that the attention is on URLs, and there is little negotiation among users in their views, thus stalemate towards the end.

The analysis of these two threads show the significance of URLs in some user-user interactions, and the intertextual functions of URLs that widen the dialogic space such that users are exposed to different voices. Some users seen to have visited the URLs posted by others, else they

would not be able to critique them. Although users seem to be in a link war in their stance-taking against each other, URLs are co-constructed by the users as evidence, and the weight depends on the credibility of the experts mentioned in the URLs. However, this singular focus on URLs may hinder intersubjectivity as illustrated in Chapter 8 where users negotiate common ground and understanding based on what each other has said, rather than the URLs. This might explain why the link wars found in these two threads end in criticism of each other's URLs and stalemate, rather than a dialogic conversation.

9.5.3 Superiority of URLs

This section explores the scenario where one user employs URLs as evidence for their stance-taking while another user does not. In these interactions, the provision of URLs becomes the point of contention. This exploration extends the quantitative findings in section 9.4.2 that found individual differences in URL-posting by investigating users' meta-pragmatic discussion on the need of URLs in online discussions. It is found that some users tend to idealize URLs as necessary evidence when disagreement arise. Three threads are exemplified to illustrate differences between users who use URLs as evidence and users who do not.

9.5.3.1 "Can you back that up with some *links*?"

This is a short thread where a user first posts a YouTube URL in response to a discussion prompt, which is challenged by another user (Figure 9.19). Interestingly, a third user asks the disagreeing user for URLs, suggesting that some users see URLs as crucial for voicing disagreement. This thread also illustrates how the relevance of the URLs posted may be misinterpreted by others if the URL is posted with minimal wording and is not elaborated on.

Figure 9.19 Thread 6400817.

Thread 6400817

Source: <https://www.futurelearn.com/courses/soils/1/comments/6400817>

Discussion prompt: “There are many other historic examples of the severe impact that soil misuse can have on society, can you identify any? What are your thoughts on these examples?”

Initiating post 2015-07-26 19:27:29 Likes: 0

User s1-1168

The Irish Famine is such an example.

<https://www.youtube.com/watch?v=5uNMGzSL42U>

Reply 1 (first contribution) 2015-07-27 00:08:16 Likes: 1

User s1-1981

The Irish famine had nothing to do with soil erosion, and in fact the link that you have posted is pure BBC propaganda and mainly incorrect.

Reply 2 (first contribution) 2015-07-27 12:05:34 Likes: 0

User s1-620

[s1-1981], can you back that up with some links?

Reply 3 (initiator’s subsequent contribution) 2015-07-27 15:16:25 Likes: 1

User s1- 1168

[s1-1981], the question we were asked related to soil misuse - not just soil erosion. The clip which I posted gives a balanced and impartial account of events. The potato blight took such a strong hold because the same crop had been planted year after year even in places previously infected. I would class that as soil misuse.

Reply 4 (subsequent contribution) 2015-07-31 12:38:46 Likes: 0

User s1-1981

[s1-1168], it is far more complicated than just growing one crop year on year, there was no shortage of food in Ireland during the famine but it was sold to the highest bidder, England. Those who died during this period had the poorest land in which to produce a crop often it was just .5 of an acre, it didn't help that there was only one variety of potato available to the poor, the Lumper, which is to this day highly susceptible to blight (and not worth eating in any case) On poor land the potato was the only crop that would sustain a family often as many as eight people. Blight is a wind borne virus, it effected all of Europe but the Irish poor were dependant on it, the grain crops were for sale to the rich in Dublin and the Irish masters in England.

Note. User s1-620 asks user s1-1981 to back up their disagreement with “links”, yet user s1-1981 writes their personal opinion and knowledge on the issue in reply 4.

In the initiating post, user s1-1168 posts a YouTube URL which is a 49-minute BBC documentary, and introduces it with “The Irish Famine is such an example”, without other writing. This is in response to the discussion prompt that asks users to identify “historic examples of the severe impact that soil misuse can have on society”. However, the user does not write their own elaboration or what has been said in the YouTube video posted, despite the discussion prompt also asking, “What are your thoughts on these examples?” It is possible that the URL is posted to substitute one’s voice in response to the discussion prompt. Another possibility is that the user’s writing “The Irish Famine is such an example” indicates their claim, while the URL serves as a supporting evidence, although there is no explanation as to how it supports the claim.

However, this URL-posting with minimal wording is bluntly rejected by another user s1-1981 in reply 1 “The Irish famine had nothing to do with soil erosion” and the URL is denounced as “pure BBC propaganda”. This rejection suggests that user s1-1981 interprets user s1-1168 as saying the Irish famine is soil erosion, which in turn is clarified by the user s1-1168 in reply 3 “I would class that as soil misuse.” The misinterpretation can be due to the lack of elaboration in the initiating post because only “such an example” is mentioned but without referring to any concept. This attests to the possibility that other users may interpret the URL differently and focus on the URL only when it is introduced with minimal wording.

Fortunately, in this case, the URL-posting user s1-1168 comes back to the thread to explain its relevance to the discussion prompt in reply 3, after being challenged. This explanation in turn prompts user s1-1981 to provide their account of their stance, instead of simply denouncing the URL as they do in reply 1. It is worth noticing that the URL-posting user s1-1168 also elaborates their interpretation of the content in the URL, rather than focusing on defending the URL per se, and user s1-1981 also elaborates on their stance, instead of focusing on criticizing the URL in their subsequent reply in the thread, thus they still engage in negotiation regarding the issue, as shown in reply 3 and 4. This contrast with the preceding section where users only focus on the URLs that deter them from discussing the content.

Nonetheless, before this negotiation happens, the URL is the point of contention in the exchange between the two users and another user. For user s1-1981, it “is pure BBC propaganda and mainly incorrect”, as mentioned in reply 1, whereas for user s1-1168, it “gives a balanced and impartial account of events”, as mentioned in reply 3. The former construes the URL as not showing fact whereas the latter attributes authoritative status to the URL. This difference points to the different ways in which both users provide evidence for their stance. As shown in the initiating post, the user s1-1168 supports their claim by posting an URL with minimal wording, suggesting the presence of an URL is the key to supporting one’s stance, without the need to elaborate on them. This contrast with user s1-1981 who explain in their own words in reply 4. Interestingly, when user s1-1981 denounces the URL as “propaganda” in reply 1, another user s1-620 asks them to “back that up with some links”, instead of asking why. This request for URL seems to suggest that URLs are needed to back up one’s dismissal of a URL posted, although some users who do not post URLs might rely on their personal knowledge in the online discussion, as have been found in this thesis and other studies (Oh et al., 2008; Sudau et al., 2014).

9.5.3.2 “Unlike you, I’m willing to give a *link*”

This thread is a conversation that is started by a question (not shown here) about the habitable zone in Mars, and drifts to a contentious issue on climate change, then a meta-discussion on evidencing practice in online discussions. The disagreement on anthropogenic effects towards climate change starts in Reply 4 and 5, followed by disputational talk in Reply 6 to 9. The provision of URL becomes the point of contention in Reply 13 to 16. The debate of climate change is not the focus of the present analysis, although there has been research on the discourse between the opposing views on this issue (e.g., Koteyko, Jaspal, & Nerlich, 2013). The focus of this analysis is on the meta-discussion among users on provision of URLs following their disagreement on climate change. The relevant parts of the thread are presented in Figure 9.20 and Figure 9.21.

Figure 9.20 First part of thread 703441

Thread 703441

Sources: <http://www.futurelearn.com/courses/moons/1/comments/703441>

[Initiating post and reply 1 to 3 are omitted]

Reply 4 (subsequent contribution) 2014-05-06 07:33:36 Likes: 1

User m1-366

The fact is that whether we humans trash our own environment or not by polluting our planet, and even filling space around it with a floating ring of junk, the real science shows that the future of the planet may be a cooler one rather than a warmer one. The Antarctic Ice cap has grown and is at a 35 year high in terms of size and thickness. According to warmist models, it should have been completely disappeared by 2013. That's just one of many examples that point in the same direction. Our Sun is weakening, and that has much more to do with global climate than so-called anthropogenic global warming. In the indeterminate future, we may find ourselves wishing that it really would get warmer.

Reply 5 (first contribution) 2014-05-06 20:26:21 Likes: 3

User m1-136

Sorry [m1-366], it is important not to overstate the importance of variations in solar output in climatic change, unreasonable as that may sound. I've done a lot of post graduate reading on the genesis of ice ages - not just those of the Pleistocene but also those going back hundreds of millions of years (snowball earth etc). In every case the most important factors are believed to be plate tectonics (particularly where it affects the clustering of continents at low or high latitudes) and variations in greenhouse gasses. Also of significance in allowing minor interglacial "holidays" (like now!), is variations in the ellipticity of the Earth's orbit and the axial tilt. For example, it is no coincidence that the Antarctic continent first acquired an ice cap 35 million years ago when the Himalayan mountains started to form. The massive peak in mountain building hugely increased chemical weathering and locked CO₂ out of the atmosphere, causing a dip in CO₂ from which we are only recently recovering. More and more research underwrites the importance of greenhouse gas variability in climate change from whatever source (volcanic, weathering, methane hydrates etc).

Reply 6 (subsequent contribution) 2014-05-07 19:59:14 Likes: 0

User m1-366

I think the jury is firmly out on that one.

Reply 7 First reply 2014-05-08 18:32:44 Likes: 0

User m1-1088

[m1-366], why reject the scientific case out of hand?

Reply 8 First reply 2014-05-08 19:14:09 Likes: 0

User m1-1671

[m1-366], what is this "real science" of which you speak? Where can I get some?

Reply 9 (subsequent contribution) 2014-05-08 21:12:10 Likes: 0

User m1-366

What scientific case? And what makes you think I would reject anything out of hand, or for that matter accept it out of hand? That's what science is all about, no?

Reply 10 (subsequent contribution) 2014-05-09 17:19:22 Likes: 0

User m1-1088

I was referring to the detail provided by [m1-136], specifically, but also to the latest reports from the IPCC and the growing consensus among scientists. I responded to your "jury's out" reaction to [m1-136]'s points.

Note. In reply 4 and reply 5, users state their stance and reasoning for and against anthropogenic climate change, but neither presents URLs.

Figure 9.21 Second part of thread 703441

Reply 11 (subsequent contribution) 2014-05-09 19:08:30 Likes: 0

User m1-136

Well said!

Reply 12 (subsequent contribution) 2014-05-09 22:17:31 Likes: 0

User m1-366

I would draw your attention to many of my other posts in which I have said that temperatures were actually higher than they are now. The jury is out because the claims of the IPCC and other groups with similar agendas are not borne out in fact. For example, even the Met Office states that there has been no warming over the last 15 years or so, despite the rise in CO₂. I personally used to believe the propaganda, until I began to look deeper into the subject, did a lot of courses in subjects such as meteorology and solar science and saw (as have done a great many highly esteemed scientists) that there are many anomalies among the claims of the GW proponents. Claims about melting polar ice, rising sea levels, and increases in hurricane activity, etc, are often simply not true. You don't have to be a genius to figure it out - you only have to look at the data compiled by NOAA, and the like. The info is right there for those who care to take an objective look at it.

Reply 13 (subsequent contribution) 2014-05-10 09:28:25 Likes: 0

User m1-1088

[m1-366], please provide some names of those highly esteemed scientists.

Reply 14 (subsequent contribution) 2014-05-10 15:18:28 Likes: 0

User m1-366

Why don't you go and look them up for yourself? Like I said, the info is out there for those who truly want to seek it. If all you're going to do is simply go into attack mode, then it probably won't do you much good and you'd be better off sticking with received opinions and what the media feeds you!

Reply 15 (subsequent contribution) 2014-05-11 11:46:39 Likes: 0

User m1-1088

[m1-366], it seems to me you're the one who is in attack mode and are making statements you won't back up. I actually have also spent a lot of time and effort investigating the climate change/ global warming question and have come to the opposite conclusion to you. Like you, I am quite highly qualified in relevant areas. Unlike you, I'm willing to give a link to enable others to start investigating:

<http://www.ipcc.ch>

Alternatively, these make interesting reading:

<http://energyblog.nationalgeographic.com/2010/12/21/climate-change-myth/>

<http://environment.nationalgeographic.com/environment/global-warming/?source=NavEnvGlobal>

Anyway, I think that you have initiated a severe deviation from the topics of this Moons course, and I do not wish to take up any more of the time and space that is supposed to be for relevant discussion. My apologies to the Course staff and students for getting drawn in to this; I just don't like hijackers going unchallenged.

Reply 16 (subsequent contribution) 2014-05-11 16:12:25 Likes: 0

User m1-366

Just a short reply to this, as I too wish not to divert the discussion topic. In my defence, it must be said that it wasn't me who began mentioning the dreaded GW. Because I was merely responding to an existing discussion, and am not here to champion any particular viewpoint, it's not up to me to provide links and references beyond those I've already cited. It seems that people have already made up their minds on the matter anyway, whatever the science says.

Note. User m1-1088 provides URLs while differentiating themselves from m1-366 who does not provide URL.

In replies 4 and 5 in Figure 9.20, both users m1-336 and m1-136 voice their opposing stance and reasoning regarding anthropogenic effects on climate change. Neither presents URLs, but indicates their bases of argument, such as “the Antarctic Ice cap” and “warmist models” in reply 4 and “post graduate reading” in reply 5. After this stance-taking, reply 6 to 11 is a disputational talk between m1-366 and four users, as shown in their short and blunt responses “the jury is firmly out on that one” and rhetorical questions (Heritage, 2002), “why reject the scientific case out of hand?”, “what is this “real science” of which you speak?”, “what makes you think I would reject”. This disputational talk then evolves into a conversation between m1-366 and m1-1088 (Figure 9.21) on provision of URLs, which is the focus of this analysis to understand users’ use of different evidence in their disagreement.

A contrast is found in the discourse practice between m1-366 and m1-1088 in their arguments and way of evidencing their claims. User m1-366 writes in relatively long comments against anthropogenic effect on climate change in reply 4 and 12, and has made some strong claims “That’s just one of many examples”, “Our Sun is weakening” and “Claims about melting polar ice, rising sea levels, and increases in hurricane activity, etc, are often simply not true”. They mention established institutes, including the Met Office and the NOAA, and criticize IPCC but do not provide any URLs. In contrast, user m1-1088 “come[s] to the opposite conclusion to” m1-366, although they never make any explicit argument, besides mentioning “the latest report from IPCC” and posting three URLs in reply 15. This difference in discourse practice underlies their meta-discussions about the provision of URLs in the online discussion.

These two users construe URLs differently, as manifested in their responses to each other. User m1-1088 seems to assume that links are essential in comments. This is evidenced in their accusation of m1-366 in reply 15, “making statements you won’t back up”, while differentiating themselves from m1-366 by suggesting that, “[u]nlike you, I’m willing to give a link to enable others to start investigating”. The accusation followed by provision of URLs indicates that URLs are construed as important to back up one’s stance by user m1-1088. However, it is interesting that m1-

1088 does not query m1-136, who writes in length in reply 5, for URLs probably because they hold the same stance, suggesting that URLs is only needed when users disagree but not when they agree, similar to the request of link in previous threads.

Despite providing URLs in reply 15, m1-1088 does not relate the URLs to climate change but seems to use them for demonstrating their “knowing” status, such that the listing of the URLs is introduced by “I am quite highly qualified in relevant areas”. This sole reliance on URLs for counter-argument further attest to the idealization of URLs in evidencing and perhaps as “esteemed” and “qualified”. The way that m1-1088 presents the URLs is thus similar to the practice of presenting unaccompanied URLs or with minimal wording discussed earlier. Nonetheless, it is possible that user m1-1088 is primarily concerned with sharing information alternative to the views of m1-366 to “enable others to start investigating”.

In contrast, user m1-366 does not provide any URLs after being queried, but repeatedly claims that one can search for information and make judgements oneself. This is evidenced in reply 12, “you only have to look at the data compiled by NOAA, and the like. The info is right there for those who care to take an objective look at it” and reply 14, “Why don't you go and look them up for yourself? Like I said, the info is out there for those who truly want to seek it”. These replies show user m1-366’s emphasis on one’s agency in distilling information, as shown in “those who care to take an objective look” and “those who truly want to seek it”. This can be further confirmed in m1-366 accusation of m1-1088, “you'd be better off sticking with received opinions and what the media feeds you!”. User m1-366 seems to construe easy-to-access URLs as secondary sources and not objective, similar to some users’ construal of Wikipedia as discussed earlier. In contrast to providing URLs, user m1-366 construes evidencing as something that requires initiative and agency. This might explain m1-366’s evidencing practice of not posting URLs but citing the institution, as shown in “look at the data compiled by NOAA”, “the Met Office states that”, “it's not up to me to provide links and references beyond those I've already cited.” Therefore, it seems that user m1-366 has a basis for their argument, although they do not provide any URL as evidence.

In short, when users have opposite opinions on a highly contentious issue, their disagreement can extend to their evidencing practices. Their discourse practices in interactions reveals their differing ideas towards URLs. In this thread, it is found that a user posting URLs may idealize URLs as important evidence, whereas a user who do not post URLs see that they must make their own decision rather than basing on “media”. This difference confirms Polletta et al.'s (2009) speculation that while some users find URLs important for supporting one’s stance, to some users online sources linked to URLs can be less authentic. Regardless of either ideology towards URLs, users focusing on provision on URLs may overshadow the chance of negotiation based on content or information, similar to the occasions where users only focus on URLs posted by each other.

9.5.3.3 “We have provided *links*”

This thread is also a conversation where disagreement over a contentious issue, homeopathy, drifts to a meta-discussion on provision of URLs as evidence in online discussions, and possibly personal attack (not discussed here). This is a long thread with 41 replies involving two disagreeing users, ah1-12 for homeopathy and does not post URL, and ah1-993 against homeopathy and posts eight URLs in this thread. Each contributes 17 replies. Six other users also contribute one to two replies and five URLs, some of them are for homeopathy and some are against.

Given that this thread is too long to be presented, only interactions between these two users that are relevant to URL-posting and evidencing are shown. Not shown here include their discussions about pharmacology, clinical trials, placebo effect, scientific methods, their identity performance, and their personal challenge to each other. This analysis focuses on the meta-discussions on URLs rather than these other issues is also driven by the observation that both users keep coming back to the issue of URLs in this thread. It is also worth noting that this analysis focuses on discourse, instead of the validity of homeopathy, similar to Koschack et al.’s (2015) examination of online forum on alternative treatments.

The two disagreeing users engage in different evidencing practices, which underlie their meta-discussion on evidencing and provision of URLs. User ah1-993, who is against homeopathy, posts and quotes eight URLs in this thread. The user employs a positive evaluative frame to introduce their URLs, by highlighting or quoting experts or authority mentioned in the URLs, as shown in Figure 9.22.

Figure 9.22 User ah1-993's replies that introduce or quote URLs

Reply 7: it is *worth* quoting "In fact, the *Australian National Health and Medical Research Council* has just published..."

Reply 13: This article is by a *homeopath*, a *Professor* of Complementary Medicine, who says 'Our trials failed to show that homeopathy is more than a placebo.....' [URL]

Reply 18: "The *World Health Organisation* (WHO) says there is no scientific evidence....." [URL]

Dr Peter Fisher (the Queen's homeopath) when he found out that homeopaths in the UK thought they could prevent malaria "I'm very angry about it....." [URL]

Reply 20: A *nice* overview of homeopathy here [URL]

Reply 22: See *Professor David Colquhoun*, a pharmacologist at University College London [URL]

Note. The emphasis is mine.

This discourse practice of introducing URLs equate the URLs with expert knowledge through their emphasis on experts and authority, along with the positive evaluative introductory frame. This is further evinced by the user ah1-993's summary towards the end of the thread in reply 24 (not shown here), where the name and credentials of the experts and authority mentioned in all the URLs posted are repeated. This frequent use and quoting of URLs within the same thread suggests the user's reliance on URLs in the online discussions. The URL-posting practice of this user also indicates an evidencing practice that count on expertise knowledge. In contrast, user ah1-12's evidencing practice is based on their own situated knowledge as a practitioner of homeopathy and "ex-medical

doctor”, as shown in Figure 9.23. The user talks about their own experience, instead of providing URLs.

Figure 9.23 User ah1-12’s replies that are about their personal experience.

Reply 8: [.....] the hundreds of patients I have helped and cured [.....]

Reply 15: [.....] Recently in NZ we had a patient dying from septicaemia in ICU; the proposed "alternative" treatment was high dose vitamin C (not homeopathy indeed) yet it needed the threat of a court case to have it done despite the fact the patient was dying...he walked out on his feet to the shame of the hospital. [.....]

Reply 16: [.....] I have seen and keep seeing the results of homeopathy and all the other non-conventional techniques I practice. Can you say the same? [.....]

Reply 21: [.....] We cure people, we do it well, we do not exploit them, we dot overcharge them and we are proud of what we, how do it and of our results.

Reply 36: I had a quick look at my last 12 months practice. Most of the "serious condition" patients I saw came AFTER they had been treated conventionally and it failed, they were left to go home, make their peace and die. Some are still alive, others had a few more pain-free months to enjoy with their families. Whether this is really due to homeopathy or not is indeed difficult to assess, and yet there is an accumulation of those patients over all my years of practice. [.....]

This contrast may explain the individual differences in posting URLs in the online discussion, where there are users who do not post URLs at all and those who share URLs in at least one-fifth of their comments. To some extent, the evidencing practice of these two users mirror the contrast between the use of expertise knowledge and situated knowledge in other online discussions and its potential conflict (Epstein, Farina, & Heidt, 2014; Koschack et al., 2015; Shanahan, 2010). User ah1-993 relies on URLs which report on expertise knowledge, whereas user ah1-12 relies on their first-hand experience in homeopathy, although ah1-12 also refers to books and foundations which promote homeopathy (not shown here). This difference in evidencing practice may hinder their interactions because neither take up each other’s evidence or engage in integration of both types of

evidence (Epstein et al., 2014; Shanahan, 2010). Instead, both users keep on contributing evidence against each other, and criticising evidence provided by each other.

This difference in evidencing practice, on top of their disagreement over homeopathy, may underlie their meta-pragmatic discussion on the provision of URLs and how they construe URLs. User ah1-993, who has been posting and quoting URLs, conceptualizes URLs as necessary evidence in online discussions, as shown in Figure 9.24.

Figure 9.24 Some of user ah1-993's replies that reveal the importance of URLs to them.

Reply 22: [.....] instead of informing us all about HOW homeopathy works or *actually providing links* to some of these wonderful studies that prove it works, your argument seems to consist of telling me I don't know what I am talking about or sending me away to spend hours (days?) reading 93 pages of references. *No links, here, on the forums, as I have done, to back up what you say.* Sadly, your argument comes across as all bluster and is not convincing.

Reply 24: [.....] So you see, it isn't just my opinion. Not only that, but we *have provided links and all this evidence is immediately available to anyone who reads this thread.* I hope that has helped other students.

Reply 29: The problem is [ah1-12] - while *I and others have been presenting actual, real, available links to evidence that backs up what we say, you have not.*

You have told me to go away and find an obscure book, or read pages of links that you can't be bothered to put on here. *Maybe they don't exist.* You are safe in making that claim and then making out it is my fault for not taking you up on the offer [.....]

Note. Emphasis is mine.

In the three replies in Figure 9.24, ah1-993 repeatedly mentions that they “have provided links” whereas ah1-12 has not. From these replies, links are conceptualized as “actual, real, available” evidence to “back up” what one says. This conceptualization of links suggests that presenting URLs is necessary in online discussions such that other users can easily access them, in contrast to “obscure book” or “pages of references” that are not posted in the discussion space. This contrast construes URL-posting as the norm of evidencing in online discussions when disagreement

occurs, and not being able to display it can be frowned upon. The challenge to ah1-12 for not displaying URLs, and emphasis on their having provided URLs, “No links, here, on the forums, as I have done”, also suggests that displaying URLs is a way to legitimize one’s knowledge and argument, and possibly construe URLs as a currency for winning an argument. This conceptualization of URLs as evidence also corresponds with the user’s discourse practices of introducing their URLs with positive adjectives and attributing the URLs to experts.

In contrast, ah1-12 construes URLs as biased sources, and other users should seek information themselves. This is evidenced from their criticism of some of the URLs posted by ah1-993, and personal challenge towards ah1-993. Their criticism towards URLs also speaks to their emphasis on personal first-hand account. Some of ah1-12’s replies that illustrate these discourses are shown in Figure 9.25.

Figure 9.25 Some of user ah1-12’s replies that reveal their negative sentiments towards URLs.

Reply 8: [.....] Knowledge starts with facts and investigating them properly, *not like the link* above or the Australian "Enquiry" that starts with a conclusion and tries to demonstrate that the conclusion is correct. BTW, that Australian enquiry has been totally debunked and shown to have multiple methodological errors and biases [.....]

Reply 19: [.....] I see that you are talking theory based on what others unilaterally say without proper checking. So much for scientific, university level approach! [.....]

Reply 27: What is fascinating is that the CLINICAL results, *the reality of the consulting room*, of the facts on the terrain are of no importance either to you, to Coquhoun (pharmacology, not clinical) to Ernst (a disgruntled and fully disregarded practitioner by the whole natural medicine community) [.....]

The replies in Figure 9.25 shows examples of how user ah1-12 critiques the URLs posted by ah1-993. It is similar to the threads shown in the previous section where users critique the ethos of experts or authors of the URLs, “starts with conclusion”, “fully disregarded practitioner”, suggesting that the user may not take up the information from the URLs, and the user construes the URLs as

“biases”. This perception of URLs is further evidenced in this user’s comments toward ah1-993 whom they regard as following “what others unilaterally say without proper checking”. User ah1-12 also highlights ah1-993 as placing no importance on “the reality of the consulting room”. Similar comments by ah1-12 regarding URLs posted and user ah1-993 can be found across the thread. These comments suggest that, in contrast to seeing URLs as “real” evidence like ah1-993 does, ah1-12 see URLs as biased secondary resources such that users should take initiative to understand first-hand information. The absolute dismissal of URLs by ah1-12, coupled with idealization of URLs by ah1-993 creates a situation where both not only stick to their stance on homeopathy, but also their own evidence, such that there is no negotiation for intersubjectivity, but personal challenge to each other. The tension between using URLs and personal account as shown in this thread also attests to Polletta et al.’s (2009) speculation that although URLs provide sources of information and evidence, some users may find URLs less authentic than personal accounts.

In summary, these three threads illustrate two evidencing practices that users employ in online discussions. The idealization of URLs as expertise evidence, and possibly final say on issues, seems to clash with personal interpretation or experience. This clash becomes obvious following a disagreement over contentious or unresolvable topics - Irish famine, homeopathy and climate change. The clash is observed through their meta-discussion on the provision of URLs in online discussion, where URLs are idealized by some as authority and expertise, but as secondary or biased by others. This meta-discussion is consistent with the analysis presented in the previous subsections which shows that some users appreciate the provision of URLs, while there are times that users question the reliability and usefulness of URLs. The differing ideals over the provision of URLs, as opposed to personal account, seem to further polarize the disagreeing users, as shown in the last two threads, rather than allowing them to see alternative views in a discussion space and possibly engage in an exchange of views. In contrast, in the first thread regarding Irish famine (Figure 9.19), after receiving criticism from another user on the credibility of the URL posted, the URL-posting user comes back to explain the relevance of their URL. Although the thread is relatively short-lived, the

users' conversation does not solely focus on the URL itself after the user explains the relevance of the URL posted.

9.5.4 Conclusions regarding micro-analysis

The three scenarios examined in the micro-analysis reveals the tensions that can arise when users post URLs linked to a wide range of online sources, and some users rely heavily on URLs while majority do not, as shown by the quantitative analysis in section 9.2. More importantly, the micro-analysis reveals that users' individual differences in employing URLs are related to their approaches to evidencing, and can sometimes trigger a link war. The resulting practices, especially when URLs themselves become the focus of attention, can hinder the processes of intersubjectivity necessary for dialogic conversations in online discussion. This conclusion is explained below while the potential of URLs generating conversations is first acknowledged.

URLs generate conversations

Micro-analysis shows that URLs can trigger sustained conversations among users. First, conversations arise when users engage in the co-construction of the value or usefulness of the URLs. This is shown by the contrasting responses to Wikipedia in different threads, in which users evaluate the relevance of the source in the immediate communicative context, rather than simply dismissing or endorsing a source. This also suggests that URL-posting in the online discussion of FutureLearn is subjected to peer-monitoring, if not by facilitators, which may challenge the spread of misinformation because users can critique a source together.

Besides, URL-posting can sometimes trigger the posting of more URLs, such that the user-user interactions can sometimes become URL-URL interactions. The URL-URL interactions can arise in two situations: when users agree with each other, such that the contribution of URLs becomes cumulative talk (Mercer, 2004); or when users disagree with each other, such that URLs are used to argue against each other in disputational talks which can be seen as link wars given the involvement of URLs, as explained next.

Link wars

Users may rely on different online sources for evidencing in their stance-taking. In particular, the micro-analysis reveals how online discussions among users can evolve into what I term *link war*, by which users privilege URLs as a necessary currency in a debate, or attribute expertise status to the URLs posted by themselves while deprecating the URLs or personal experience shared by others. This can be stemmed from users' focus on the URLs, rather than the content linked to the URLs shared. This is similar to what Wikgren (2003) finds in newsgroup, where users do not indicate the need for reading the content of the URLs posted, although they show a preference towards scientific sources over other sources linked to the URLs to legitimate their claims.

Although exposure to various kinds of sources can expand the dialogic spaces to alternative voices, users may stick to the URLs they post and dismiss the URLs posted by others. The presence of URLs, construed by some users as real and concrete evidence, may also explain why it is hard to negotiate or change internet users' stance on a certain issue when they have already come across links that support their stances (Jacobson et al., 2016; Savolainen, 2014; Wikgren, 2003).

The presence of URLs may also undercut the value of personal account in the online discussions, as speculated by Polletta et al. (2009). Users do not always accept evidence other than URLs, especially when they disagree on contentious issues. One interesting observation is that, within the same thread, users do not ask those who share the same view with them to present URLs, but only those who disagree with them, and in threads where users focus on URLs posted, replies containing no URLs might not be taken up in the conversations.

On the flip side of the link war is when users completely dismiss the URLs posted and conceptualize URLs as biased or less authentic (Polletta et al., 2009). This link war mirrors a wider ideological clash on evidencing such that there are users who rely on experts and authority while there are users who believe in situational experience and personal understanding (Bellander & Landqvist, 2020; Epstein et al., 2014; Shanahan, 2010). Although not explored in this thesis, evidencing practices may also vary across different topics (Oh et al., 2008). As shown in previous

studies on online discussions, users seem to focus on personal knowledge for investment advice (Connor, 2013), local issues (Polletta et al., 2009) and personal illness (Sudau et al., 2014), while Savolainen (2014) and Wikren (2001) find that users may rely more on expert knowledge on issues such as climate change and health.

Although both types of evidence have their value in stance-taking, this division in evidencing may hinder the intersubjectivity necessary for dialogic conversations if users hold a strong view on either type of evidence and do not recognize the possible differences underlying their evidencing practices. The ideological differences in evidencing may parallel the differences in worldview and identity that some users attribute to when they have irresolvable disagreements, as shown in the analysis of agree to disagree in Chapter 8. However, in link wars, there is no indication that users acknowledge such differences in their evidencing practices, or specify the conditions under which a certain source of evidence are relevant, thus little dialogic space for negotiation and intersubjectivity.

Although most meta-discussion about URLs analysed in this section seems to show little intersubjectivity, the meta-discussions about Wikipedia seem to show that users engage in processes of intersubjectivity. This is probably because the issue for which users share Wikipedia links is not as controversial as climate change, homeopathy, fat vs. sugar diet. Most importantly, the users posting Wikipedia links explain the source in relation to the specific context of their discussions. In contrast, in the link wars, the URLs seem to be employed to make a more broad-based argument on controversial issues, and the disagreeing users focus on the provision of URLs and authority of their own URLs, rather than the information or the relevance of the URLs in different situations, thus reducing the chance for intersubjectivity.

In short, it is hard to establish intersubjectivity for negotiation when users only focus on the URL itself or the authors or experts mentioned in the URLs. Users may miss the possible common ground in the content or argument informed by the sources linked to the URLs, or fail to see other users' personal experience or interpretation. Introducing URLs with minimal to no wordings also

deprives other users of the chance to understand the differences and similarities in their stance. The discourse practice of fixating on only one way of legitimating one's claim, for example the presence of URLs and experts, may shrink the dialogic space, because alternative voices are not considered and intersubjectivity is not negotiated.

9.6 Conclusion

The analysis in this chapter extended previous MOOC study (Gallagher & Savage, 2016) that only counts the number of URLs posted by users and established that URL-posting in MOOCs, is parallel to other online spaces examined in previous studies. Although URLs are not posted by every user, investigation of URL-posting is one way of understanding users' sharing of information in the wider online world (Polletta et al., 2009).

As with the relatively infrequent long threads, micro-analysis of users' conversations involving URLs provide a rare opportunity for researchers to explore URL-posting in online spaces. This study illustrated the underlying tensions between users who share URLs and those who do not, and users who rely on different type of sources linked to URLs, as speculated by Polletta et al. (2009), Savolainen (2014) and Wikgren (2003). The underlying tension seems to originate from the ideological differences between using situated experience and authority and expertise knowledge for evidencing, and this tension at times can hinder intersubjectivity between users. This finding thus addresses RQ3 regarding how URL-posting may sustain or hinder dialogic conversations among users in online discussions.

The micro-analysis shows that users seem to scrutinize the URLs posted, which could be good practice from the perspective of digital literacies (Gilster, 1997; Lankshear & Knobel, 2006). However, the analysis also reveals that this scrutinization can border on fixation for some users. The concrete presence of URLs supporting one's stance might be the reason why users may not explore other alternative voices but keep sharing their URLs, and might also explain the circulation of

(mis)information in online spaces in general (Jacobson et al., 2016). Overall the findings suggest that users may need to be reminded of the value of different types of sources and evidence, and the possibility of conflicting online sources that require them to examine the content and its relevance in a specific situation, besides the authority and credential of the links.

Chapter 10

Discussion and Conclusion

10.1 Introduction

This study investigated the dialogic nature of online discourse in order to raise internet users' awareness of their discourse practices in establishing social interactions and engaging in deliberation in online spaces. The online discourse was examined by focusing on users' interactions in text-based asynchronous online discussions, involving both user-user interactions and user-content interactions (Herring, 2013; Kraut et al., 2011; Ksiazek & Lessard, 2016). The former mainly occur when users reply to each other within a thread, which consists of initiating post and replies, so dialogic conversations among users are most probably seen in threads. Independent posts proved a useful point of comparison with initiating posts and were also considered because of what they might reveal about the nature of user-content interactions.

This study dissected comments in online discussions into initiating posts, independent posts, and replies to examine online discourse with a corpus linguistic approach. Specifically, keyword analysis comparing these three types of comments revealed lexical differences which in turn pointed to distinct discourse practices among them. This analysis thus addressed the first two research questions:

RQ1: What are the differences in the linguistic features and discourse practices that regularly occur in

- **initiating posts that receive replies and start a discussion thread,**
- **independent posts that do not receive replies,**
- **replies, especially those in sustained discussions**

RQ2: How do these discourse practices initiate, sustain or hinder dialogic conversations in the online discussions?

This study further explored online discourse by zoning into one discourse practice specific to online spaces – URL-posting – and addressing the third research question:

RQ3: How does URL-posting initiate, sustain or hinder dialogic conversations in the online discussions?

This corpus analysis into the discourse practices in text-based asynchronous online discussions contributes to existing field of knowledge in three ways: empirically, it shows that users execute their agency via language to initiate and sustain conversations with others, and to employ information sources online; theoretically, it adds to a description of the dialogic nature of online discourse and distinguishes dialogic conversations from user-user interactions; and methodologically, it combines corpus linguistics with micro-analysis that effectively investigate both general patterns and nuanced discourse practices in big language data available from MOOCs.

In this final chapter, the findings are first summarized, followed by the empirical, theoretical and methodological contributions of this thesis. Practical implications in digital literacies are also discussed. Lastly, limitations of this study are acknowledged and future research suggested.

10.2 Key findings

In this thesis, five key findings regarding online discourse and users' interactions in online discussions can be concluded, and are elaborated on in the following subsections:

- (1) Users' interactions in the online discussions are observed to be prompt-focused which is mediated by platform design, attesting to the nature of discourse in internet-mediated communication;
- (2) At the start of user-user interactions, a dialogic space and a potential relationship with others are constructed by users' discourse, showing that the concept of dialogic space is useful to explain the dialogic nature of online discourse;
- (3) Disagreement can trigger sustained user-user interactions that provide opportunity for achieving intersubjectivity through negotiation, highlighting the role of disagreement in online discussions, and the potential of MOOC discussions, specifically FutureLearn, in avoiding echo chambers by allowing users who hold different points of views to engage with each other;
- (4) Certain discourse practices are double-edged swords that can either shun or expand dialogic space in discussion threads, demonstrating the need for users to be aware of best discourse practices in online discussions to achieve intersubjectivity and thus to engage in dialogic conversations;
- (5) URL hyperlinking, emerged from the inductive keyword analysis, reflecting not only the information exchange culture in the online discussions, but also evidencing practices among users with opposing stances that can at times hinder the processes of intersubjectivity.

10.2.1 Prompt-focused posting

The large number of independent posts, and short-lived discussion threads observed in the Futurelearn MOOC online discussions mirrors the prompt-focused posting in other online spaces (Herring, 2013; Ksiazek & Lessard, 2016; Beth et al., 2015; Burke, et al., 2007; Marcoccia, 2004; Meyer et al., 2019). This prompt-focused posting is also evinced by the observation in this thesis that most users create more posts than replies and seldom engage in subsequent contributions in the

threads that they have initiated or replied to before. One-time contributors and one-reply threads are also common. However, the FutureLearn MOOC online discussions also contain exceptional cases of super-posters and sustained threads, as in other online spaces (Huang et al., 2014; Lambiase, 2010; Poquet et al., 2018; Stommel & Koole, 2010). This suggests the complexity and heterogeneity of interactions within online discussions.

As pointed out by Herring (2013), although online space is characterized by both user-generated content and social interactions, the former usually dominates. Users tend to comment in response to the initial prompt on the web, such as news story and video, rather than in response to others' comments. This is exactly what has been observed in the FutureLearn online discussions where there are more independent posts, and the discourse of these posts reveals users are in a conversation with the content or the course designers. To some extent, this can be considered as a success to the host of the web because part of the FutureLearn design is to encourage users to have a reflective conversation within themselves along their learning journey (R. Ferguson & Sharples, 2014; Preece & Maloney-Krichmar, 2005). However, this prompt-focused posting could have a double-whammy effect on the lack of replies to these posts because users may be overwhelmed by the large number of independent posts in the online discussion spaces such that they do not reply to others' posts (Himmelboim, 2008).

The prompt-focused posting may also partly explain the large number of one-reply threads and short threads. Users may contribute once in a thread, either by initiating the thread or replying once in the thread, and never come back to the thread again. Fortunately, some users, especially super-posters, continue their engagement in a thread by making subsequent contributions such that sustained interactions beneficial for intersubjectivity still occur amid independent posts and short-lived interactions. Furthermore, the polylogal nature of online discussions also allows other users to join in the same thread to continue a discussion left off by those who do not come back. This has been observed in the current analysis where users whose initiating post starts sustained interactions among other users but they themselves never come back to the thread.

More importantly, the prompt-focused posting and the polylogal nature of online discussions reflect the fact that online spaces create a leveling ground for users to voice out without the need to fight for conversational floor or to respond to interlocutors as typically required in face-to-face interactions (Baym, 1996; Cavanagh, 2007). Although this attests to the user-content interactions, Bou-Franch & Garcés-Conejos Blitvich (2014) also argued that this can be an indication that users are more interested in expressing their views rather than engaging others' views in a discussion, thus compromising the user-user interactions which some other users value (Delahunty, 2018; Hew & Cheung, 2014; Hewings et al., 2009; Joyce & Kraut, 2006; Springer et al., 2015). Furthermore, posting and replying without entertaining others' voices is rather monologic than dialogic and is not constructive for online deliberation (Dahlberg, 2001; Freelon, 2015; Friess & Eilders, 2015). Nonetheless, as shown in the corpus analysis in this thesis, there are several discourse practices facilitative of establishing and expanding dialogic space for user-user interactions, attesting to the agency that users can potentially take if they would like to initiate social interactions amid the prompt-focused posting phenomenon in online spaces.

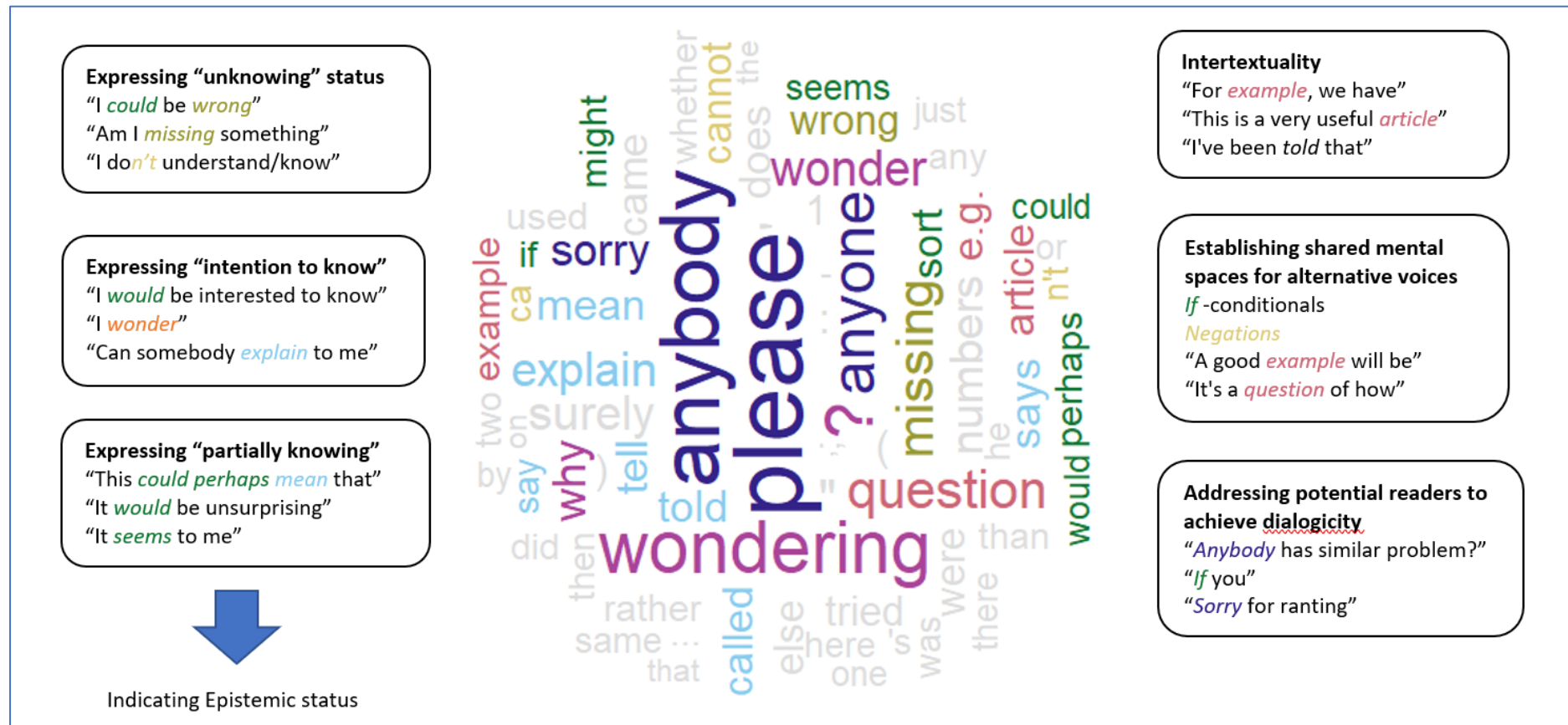
10.2.2 Discourse creates relationship among users in a dialogic space

The lexical differences found between initiating posts and independent posts shows that discourse practices create different types of relationships in online discussions, confirming the role of language in construing relationships in social world (e.g., Bakhtin, 1981; Fairclough, 2003; Heritage, 2012; Herring, 2004; Vygotsky, 1978; Martin & White, 2005). In the initiating posts that start a thread, discourse establishing a prospective dialogic relationship to real or imaginary audience renders a post more likely to receive a reply. This relationship could also be an epistemic one in which users construct the initiating posts by indicating their "unknowing", "partially knowing" or "intention to know" status to invite others who might be at a better "knowing" status to fill the gap, whether for information seeking or stance-taking (Concannon et al., 2017; Heritage, 2012).

Besides creating an epistemic or dialogic relationship with other users, the discourse found in the initiating posts also attests to the role of language in constructing a dialogic space for multiple voices (Delahunty, 2018; Du Bois & Kärkkäinen, 2012; Lapadat, 2007; Mercer, 2004; Stahl, 2015; Wegerif, 2010). The discourse in initiating posts that has the function of intertextuality, or use *if*-conditionals or *example* to create hypothetical, counterfactual situations or real-life examples, provides a concrete common ground for others to comment on and develop their interactions (G. Ferguson, 2001; Landqvist, 2016; Liu & Liu, 2017). Negative propositions may also invoke the positive counterpart, thus setting up a dialogic space of multiple views. Similarly, the qualification of one's stance with *if*-conditionals, modals or hedges also allows alternative voices to pitch in (Martin & White, 2005). A dialogic space is thus created or expanded via these discourses which indicate the possibility of multiple voices such that others are welcomed to join the conversations.

This finding thus reveals the discourse practice that regularly occur in the initiating posts that start a discussion thread. These discourse practices can potentially initiate dialogic conversations by creating a dialogic space welcoming and inviting others' voice. This finding is summarized in Figure 10.1.

Figure 10.1 Word cloud of initiating keywords



Note. The font size indicates relatively the effect size, i.e., the bigger the font, the more often the keyword is used in the initiating posts compared to the independent posts. Keywords with the same colour share similar prevalent functions, although each keyword can have multiple functions and it depends on the context they occur. The keywords in light grey are those not analysed in this thesis. The bubble summarizes the discourse practices identified in this thesis with selected examples.

In contrast, in the independent posts that do not start a thread, the discourse establishes a retrospective dialogic relationship with the course content or discussion prompts that come before. The response to the discussion prompts is equivalent to answering a question raised by the educators of the MOOCs, such that other users may not join this conversation, and may choose to answer the discussion prompt themselves by creating another new post. This is the prompt-focused posting as mentioned earlier in section 10.2.1. Therefore, the dialogic relationship established in the independent posts is one of user-content interactions rather than user-user interactions found in the initiating posts. The self-references found in the discourse of independent posts may also indicate users' reflection on their learning, which is inward-oriented rather than outward-oriented to other users. In FutureLearn terminology, it is the internal conversation users have within themselves along their learning journey (R. Ferguson & Sharples, 2014; Laurillard, 2012). Although also dialogic in nature, some independent posts are expressions of appreciation towards facilitators and educators such that other users are not the potential audiences. Epistemic status is also indicated in the independent posts, but it is an expression of "becoming knowing" such that other users are left with no gap to fill, so there may not be an epistemic relationship established with other users in this discourse.

This finding thus reveals the discourse practices that regularly occur in the independent posts that do not receive a reply. These discourse practices are less likely to initiate user-user interactions because the dialogic relationship they create is retrospective or inward-looking and the discourse indicating "becoming knowing" or self-references do not necessarily leave open a dialogic space that welcome others' voice. This finding is summarized in Figure 10.2.

[illegible]

367 | Page

At this point, it is worth revisiting the three posts presented at the start of this thesis to illustrate how the findings attest to the potential role language plays in initiating user-user interactions. These three posts are drawn from the same step in the finance-1 MOOC, which is without any discussion prompt.

Post 1 (posted on 9 April 2015, liked by 8 users)

“was bemused by the cartoon at the beinning...ok so you may not like what i write but noticed the 'rich' family had one child and the 'poor' had two...is this part of the inequality in society that some and i repeat some people have children they cannot afford but expect someone to pick up the tab by having more benefits such as tax credits child benefit needing larger houses etc ..it is just a question..family and friends who have more children are generally poorer”

Post 2 (posted on 8 April 2015, liked by 3 users)

“Should we be looking at the whole system rather than blaming the baby boomers for everything?
<http://www.theguardian.com/commentisfree/2015/apr/08/rising-inequality-technological-change-loss-jobs>
Professor Anthony Atkinson has a lot of good points - and the current system might - probably was - designed to produce, preserve and increase inequality. And there is an election on 7th May in the UK.”

Post 3 (posted on 17 April 2015, liked by 8 users)

“Have only managed to start week 4 today, Saturday, but am really saddened by some of this discussion. I've always been proud to be a taxpayer, proud that these are my roads, my hospitals, my teaching staff etc etc ... and also proud that I / we can support those who need it.”

It may now seem obvious that Post 1 and 2 are initiating posts whereas Post 3 is independent post. In Post 1, although without initiating keywords of a dialogic nature, the user addresses potential audience with “ok so you may not like what i write”, thus establishing a dialogic relationship with other users, possibly those that disagree with them. Similarly, although no keywords for intertextuality is found, “the cartoon at the beinning ...” links intertextually to the

content in the step of the course, thus setting up a common ground for the dialogic space to develop. This dialogic space is also established towards the end of the initiating post, where the meta-language initiating keyword *question* is used in “it is just a *question*” to introduce their stance “family and friends who have more children are generally poorer”. This post also invokes the opposing viewpoint regarding this stance, thus inviting users with alternative views to join the dialogic space. Admittedly, the issue mentioned in this initiating post can be contentious to start with, and this thread consists of 51 replies, making it the longest thread in the FL corpus. Post 2 only receives one reply which can be said to be in response to the question raised by the user “Should we be looking at the whole system rather than blaming the baby boomers for everything?” This question, although rhetorical in nature given the user provides a URL to answer it themselves, establishes an epistemic relationship with users, whether it is for seeking information or taking a stance. The URL-posting practice also serves to expand the dialogic space as the URL is linked intertextually to other sources. The URL-posting practice will be explained in section 10.2.5.

In contrast, Post 3, although it was liked by eight users, i.e., read by others, does not receive any reply. The first-person pronouns *I*, *my* are independent keywords, and indeed this post is a self-reference regarding the user’s learning journey “only managed to start week 4 today”, reflection “proud to be a taxpayer” and user-content interaction as shown in “really saddened by some of this discussion”. This discourse indicates that the user read other comments but wrote a reflection as a new post rather than replying to a specific post. The discourse in this independent post can thus be said to form a dialogic relationship with the content of the discussion in general, rather than with potential audiences, and there is no dialogic space for other users to pitch in.

In short, based on the keyword analysis between initiating posts and independent posts, it can be concluded that discourse that starts user-user interactions is that which creates a relationship with potential audiences who are then invited to a dialogic space, whether their voices align or disalign with stances in the initiating posts. On the contrary, discourse practices that are mainly self-references or indicates a dialogic relationship with educators or course content are less likely to

invite replies. These findings regarding initiating posts and independent posts therefore answer RQ1 and 2 in relation to the differences in the linguistics features and discourse practices that regularly occur in these two types of posts and how these practices initiate dialogic conversations in online discussions.

10.2.3 Disagreement triggers sustained interactions as a dialogic space

Although the reply keywords reveal that users generally engage in interactive discourse, addressing each other, referring to others' initiating post or reply, aligning or dis-aligning with others' stance, these interactions tend to be short-lived. In the present corpus, half of the threads are one-reply long, and another 40% two to four replies. The short-lived interactions may be due to the fact that the stance-taking is mainly agreement, such that users agree with the same stance and do not develop or explore other voices further, similar to the cumulative talk and supportive interactions found in other online discussions in learning settings (e.g., Kellogg et al., 2014; Lapadat, 2007; Littleton & Whitelock, 2005; Paulus, 2006; Rourke & Kanuka, 2007). Agreement or alignment is also found frequently in the independent posts that do not receive replies.

In contrast, disagreement, although relatively fewer, is found to be one potential trigger for the sustained interactions in the FutureLearn MOOC online discussions, attesting to the importance of disagreement in online deliberation as well as peer discussions in learning (Lapadat, 2007; Laurillard, 2012; Littleton & Whitelock, 2005; Mercer, 2004; Dahlberg, 2001; Lewiński, 2013; S. Wright & Street, 2007). This is mainly because disagreement indicates the presence of alternative voices, thus creating a dialogic space that allows exploration of different voices. However, users need to be linguistically competent to engage in a disagreement such that the disagreement does not lead to disputational talk but to exploratory talk and dialogic conversations (Baym, 1996; Bou-Franch & Garcés-Conejos Blitvich, 2014; Chiu, 2008; Concannon & Healey, 2015; Felton et al., 2015; Marra, 2012; Mercer, 2004). As will be elaborated next, this thesis shows that users' discourse practices shape how a disagreement can develop into a sustained interaction which in turn becomes

a dialogic space for processes of intersubjectivity. The practical implication of the presence of disagreement in FutureLearn MOOCs will be further elaborated on in section 10.6.2.

10.2.4 Discourse practices as double-edged swords for engaging in intersubjectivity

The micro-analysis of threads where disagreements arise identified several discourse practices

facilitative of intersubjectivity: concession and assertion, qualification with *if*-conditionals, meta-language, identity performance. These discourse practices are derived from the reply keywords. All these practices involve acknowledging others' stance while updating or maintaining one's own stance, thus creating a dialogic relationship between each subjectivity; that is, intersubjectivity. For example, in concession and assertion, users explicitly acknowledge others' point of view while maintaining or revising their own view, not only creating coherence but also integrating similarities and pointing out differences in their views. Similarly, *if*-conditionals are used to introduce different conditions under which different voices are legitimate, thus allowing exploration and integration of different stances.

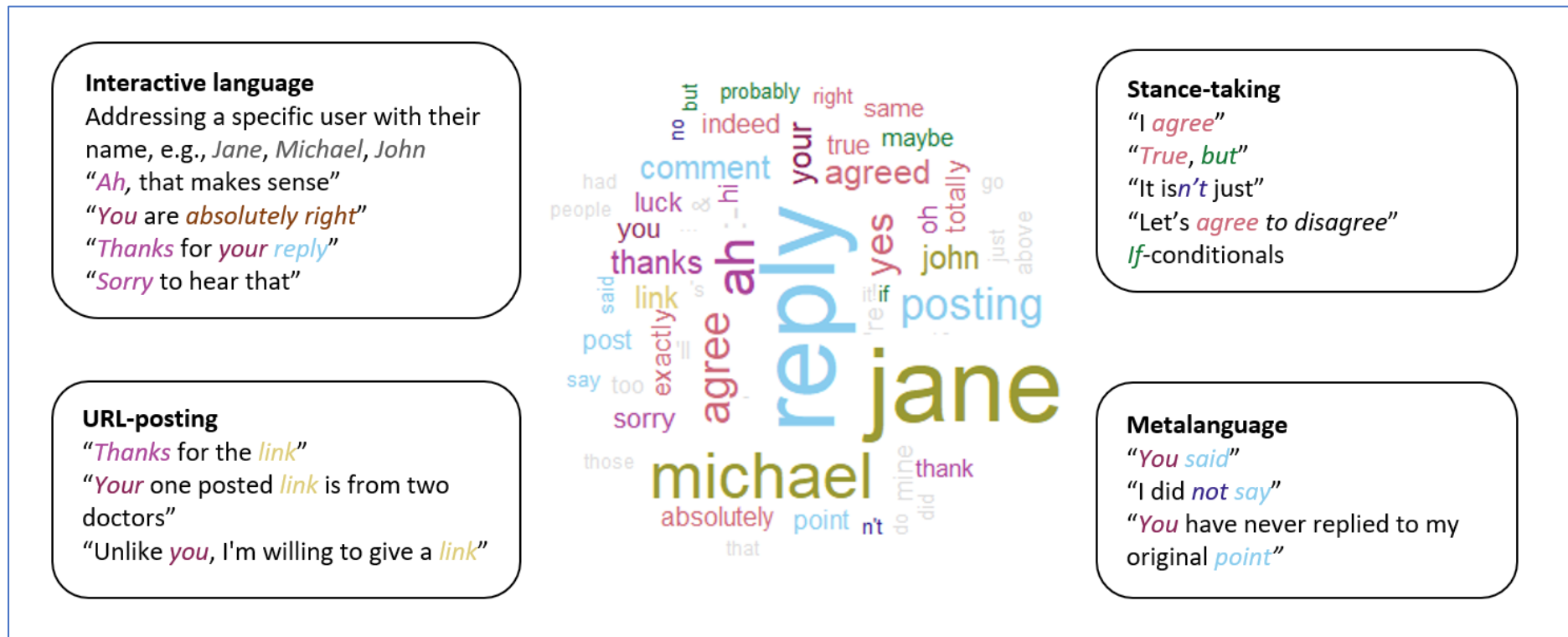
The other two discourse practices, meta-language and identity performance, however, can be a double-edged sword in shaping the negotiation of disagreement in online discussions. On the one hand, both practices make explicit comments or references to the users' own or the others' discourse or identity. Firstly, it is found that metalinguistic discourse clarifies what a user has said before or queries others' previous comments to achieve mutual understanding or to identify common ground, similar to mediation process (Janier & Reed, 2017; Liu & Liu, 2017). Secondly, identity performance is found to be used to explain their own stance or to appreciate others' stance, as in other online discussions (Bou-Franch & Garcés-Conejos Blitvich, 2014; Grabill & Pigg, 2012; Jaworska, 2018).

On the other hand, however, both meta-language and identity work can be a hindrance to intersubjectivity if the users do not take up the clarification or recognize the possibility of opposing stances. This happens when the interaction drifts into metapragmatic discussions of how others

should comment in the online discussions, or continuous metalinguistic discourse critiquing others' discourse or posting of URLs rather than what is being said, or when users position their own identity as higher than the others in terms of epistemic status, or border on attacking others' identity. The dialogic space for intersubjectivity is thus shunned in these situations despite sustained interactions because users' discourse does not acknowledge others' stance nor recognize the possibility of agreeing to disagree, i.e., that alternative voices exist such that there is no "winner" in the discussion (Felton et al., 2015; Nathan et al., 2007; Sarewitz, 2011). Because there is little to no intersubjectivity in these conversations despite being sustained interactions, these conversations can be parallel monologues rather than dialogic conversations.

In short, the keyword analysis of replies reveals that the discourse practices in replies are interactive and responsive to the initiating post or other replies in a thread. The micro-analysis of threads illustrates how users employ discourse practices that are facilitative of intersubjectivity and exploration of alternative voices for stance-taking. Most importantly, this finding demonstrates that the potential of a disagreement developing into a dialogic space for intersubjectivity is largely shaped by users' discourse, further illustrating the role of language and users' agency in online discussions. Specifically, this thesis highlights the importance of acknowledging others' views or underlying differences despite being in a disagreement, while fixation on critiquing each other's discourse, identity, or URL provision (to be elaborated in the next section) can shrink dialogic space and hinder intersubjectivity. This finding is summarized in Figure 10.3, and answer the RQ1 and 2 in relation to the differences in linguistic features and discourse practices that regularly occur in replies in comparison to the two types of posts, and how these practices sustain or hinder the development of dialogic conversations in online discussions.

Figure 10.3 Word cloud of reply keywords.



Note. The font size indicates relatively the effect size, i.e., the bigger the font, the more often the keyword is used in the replies compared to the initiating posts and independent posts. Keywords with the same colour share similar functions, although each keyword can have multiple functions and depends on the context they occur. The keywords in light grey are those not analysed in-depth in this thesis.

10.2.5 URL-posting reflects users' evidencing practice

Although not all users post URLs, this study found that hyperlinked resources are another meaning-making tool besides language in online discussions, to the extent that some users only post URLs with few to no words of their own or fixate on the presence of the URLs in their stance-taking with others. For users who rely on URLs for stance-taking, URLs are construed as concrete evidence that others can assess easily in online discussions and are attributed authoritative value over personal experience or interpretation. This idealization of URLs is observed from users' discourse focusing on the necessity of URLs and highlighting the experts mentioned in the URLs in their comments, or when users mainly include the text from the URL in their comments instead of their own interpretations.

As with discourse practices such as meta-language and identity performances, URLs can be a double-edged sword in online discussions. On one hand, URL-posting reflects information exchange among users, and generally welcomed by users, as revealed by the positive sentiments expressed towards the URLs posted. Furthermore, users are found to be engaging in a co-constructive process discussing about the relevance and value of URLs posted, suggesting peer-monitoring in online discussions and URLs can trigger interactions among users. In these occasions, URLs expand the dialogic space to voices outside of the immediate communicative context, and URLs can become a shared common ground that trigger interactions among users (Himmelboim et al., 2009). The evaluation of the relevance of the sources linked to the URLs also reflects users' digital literacies in information use (Lankshear & Knobel, 2006).

On the other hand, URLs can become a barrier to intersubjectivity between users holding opposing stance on contentious issues when both parties are able to present URLs supporting their stances, or when only one party presents URLs but the other do not. This is when disagreement evolves into a link war rather than into a dialogic space for intersubjectivity. Firstly, when both opposing parties use the mere presence of URLs as evidence, the negotiation can become a situation of "he said, she said" such that each user only focuses on the URLs themselves, instead of the

content of the URLs that might form the potential common ground for intersubjectivity (Jacobson et al., 2016). Secondly, link wars also arise when the party who uses URLs as evidence for their stance dismisses or does not take up the other types of evidence, such as personal account, presented by other users. The focus on the provision of URLs to the extent of marginalizing other types of evidence shrinks the dialogic space because other voices are not considered. Furthermore, without acknowledging the underlying difference in their evidencing practices (i.e., concrete evidence related to expertise knowledge vs. personal situated experience) the disagreement may stay at stalemate rather than developing into the processes of intersubjectivity for dialogic conversations.

In short, it emerged from this thesis that beside linguistic resources, URLs are an additional resource at users' disposal in online discussions, given the hyperlinking function afforded by the technology (Kiernan, 2018; Tyrkko, 2010). URLs posted may form a common ground for users to engage with each other, depending on their discourse practices. URL-posting also reflects users' evidencing practice in their stance-taking with others in online discussions. Specifically, some users may over-rely on easily-accessed evidence and authoritative expertise in their stance-taking, while there are users who rely on other types of evidence (Oh et al., 2008; Polletta et al., 2009; Savolainen, 2014). This difference in evidencing practices might also explain why some disagreement does not develop into a dialogic space of intersubjectivity and why users remain polarized (Koschack et al., 2015). This finding thus addresses RQ3 by showing how URL-posting initiates, sustains and hinders dialogic conversations in the online discussions.

10.2.6 Summary

The findings in this thesis show that discourse practices are employed by users in different types of comments to differential effects in establishing conversations with others in online discussions. Similarly, URL-posting, depending on users' discourse practices, also has differential effects on users' interactions. The key findings are that discourse practices that anticipate and entertain alternative voices, address potential audience and recognize differences are facilitative of users establishing

social interactions and engaging in deliberation in online spaces. In contrast, discourse practices that exclude or ignore alternative voices and fixate on one option may hinder social interactions and reduce the chance of deliberation. There are also discourse practices that function to engage in user-content interactions such as addressing the discussion prompt and course content.

10.3 Empirical contributions

As observed in previous studies of other online spaces (Beth et al., 2015; Burke et al., 2007; Marcoccia, 2004; Meyer et al., 2019), the current research setting, FutureLearn MOOC online discussions, also contained relatively more independent posts and short-lived interactions that can compromise the experience of information exchange, socialization and online deliberation. By unravelling discourse practices that users can employ to establish and engage in conversations with others (as summarized above), this thesis illustrated empirically users' agency as enacted via language in shaping their own and others' experience in online discussions.

Furthermore, the empirical investigation in this thesis goes beyond previous studies in four ways. Firstly, Chapter 5 revealed that users' posting patterns in online discussions are more complicated than suggested by previous studies which identify super-posters and one-time contributors based only on users' frequency of posting (Huang et al., 2014; Ruiz et al., 2011). Although users are generally prompt-focused, they differ in their engagement in continued interactions within a thread. Secondly, this thesis extended previous studies that only examine linguistic features by investigating how they are used in context and in discourse practices (e.g., Arguello et al., 2006; Crook et al., 2016; Chen et al., 2020). Chapter 6 and 7 not only examined the initiating posts but also the independent posts to gain a more thorough understanding of the discourse practices that can potentially start a conversation. At the same time, the function of independent posts in the online discussions, that is user-content interactions, was also revealed.

Thirdly, informed by the linguistic features in replies, Chapter 8 illustrated how conversations unfold in threads through several discourse practices that can facilitate intersubjectivity during stance-taking, on top of the expression of agreement and disagreement that have been the focus of previous studies (Baym, 1996; Kleinke, 2010). Lastly, building on previous studies that examine the sources linked to URLs posted in online discussion (Polletta, Chen, & Anderson, 2009; Savolainen, 2014; Sudau et al., 2014), Chapter 9 further illustrated that users' approaches to evidence underlie tensions in the use of different online sources and personal accounts. All these empirical findings further contribute to our understanding of both general patterns and nuances of users' posting behaviours and discourse in online spaces, while attest to the complexity of online discussion that is reflected from users' discourse.

10.4 Theoretical contributions

The theoretical contributions of this thesis are twofold: (1) the application of the concepts of dialogic space and intersubjectivity in online discussions expands our understanding of these two concepts to online discourse; (2) the findings provide a detailed empirical account of the concepts of dialogic space and intersubjectivity.

10.4.1 Expanding the concepts of dialogic space and intersubjectivity

The concepts of dialogic space and intersubjectivity have been developed to account for the dialogic nature of human language (Du Bois, 2007; Martin & White, 2005) and peer interactions in education (Mercer, 2004; Stahl, 2015; Wegerif, 2010). In this thesis, these two concepts are applied together to the online discussions of MOOCs to explain discourse practices in users' interactions, thus expanding our understanding of both concepts to online discourse.

Originally proposed by Martin & White (2005), a dialogic space is created by an utterance, and can be expanded or contracted, depending on the discourse in the utterance. As shown in this thesis, the conceptualization of the dialogic space can be extended to online discourse, such that a

dialogic space can be created by an initiating post at the start of a thread, while the effect of the dialogic expansion or contraction can be seen in the development of the threads, that is the replies. Furthermore, according to Martin & White (2005) and Wegerif (2010), a dialogic space is discursively expanded by entertaining multiple voices, while in Mercer's (2004) terminology it is an exploratory talk. As shown in this thesis, the exploration of multiple voices is not only important for inviting others' reply, whether for information exchange or socialization, but also for online deliberation between users within a thread (Dahlberg, 2001). Therefore, in online discourse, the concept of dialogic space can also be understood as a space for deliberation, as a contrast to echo chambers which normally consists of one voice (Freelon, 2015; Veletsianos et al., 2018; Walter et al., 2018).

The relationship between various voices in a dialogic space can be further specified. Common relationships include dialogic, in which previous voices are taken up or potential voices are anticipated (Bakhtin, 1981), and intertextuality, in which other utterances or social-cultural context are referred to (Fairclough, 2003). While these two relationships can be easily conceptualized within a thread, that is user-user interactions, our understanding of these relationships can be further extended to user-content interactions in online spaces. It was shown in this thesis that users can establish dialogic relationship and intertextuality to the content on the webpage or authors of the page, in this case educators who design the course.

Besides dialogic and intertextuality, there are other possible relationships between voices within a dialogic space in online discussions. In this thesis, intersubjectivity is introduced to further explain the relationships between these voices, especially within the threads. These relationships include epistemic (Heritage, 2012), (re-)calibration and (dis)alignment of stance (Du Bois, 2007). Importantly, in this thesis, stance-taking has clearly characterized online deliberation where users not only raise various voices, but also integrate alternative voices into their own voices, as evidenced by their discourse practices summarized above. Furthermore, the findings in this thesis also shows that a continuous (re-)calibration of the interrelationship between the different stances is useful to understand the situations where users remain polarized in online discussions. It shows that the

outcome of negotiation does not need to be converging to agreement but recognition of each other's stance and possible re-calibration of one's own stance. Therefore, the introduction of intersubjectivity into the concept of dialogic space further extend our understanding of dialogic space in online spaces.

Finally, as have been argued by numerous researchers (Delahunty, 2018; Freelon, 2015; Hall, 2010), simply voicing opinions and user-user interactions may not be constructive for online deliberation or peer discussions. This is especially possible in text-based asynchronous online discussions where users are free to express themselves but are not obliged to read or take up others' responses (Cavanagh, 2007; Herring, 2013). It is possible a user-user interaction can be parallel monologues, in which although multiple voices are raised in the dialogic space, they are not necessarily entertained or in a relationship as specified above. This thesis successfully distinguishes dialogic conversations among the threads in the online discussions. Dialogic conversations can be characterized by integrating dialogic space and intersubjectivity, thus contributing to the conceptualization of user-user interactions that are constructive for online deliberation or when users holding onto opposing stances. Therefore, this thesis not only extends the concepts of dialogic space and intersubjectivity to online discourse, but also contribute to our understanding of user-user interactions in online discussions.

10.4.2 Empirical accounts of dialogic space and intersubjectivity

This thesis expanded the boundary of dialogic space and intersubjectivity to include online discourse. It showed that a dialogic space can be co-constructed by users through certain discourse practices and, importantly, how this discursively created dialogic space is key to attracting replies and sustaining interactions in online discussions. The analysis successfully enriches the concept of dialogic space with an empirical description of discourse practices that make dialogic space possible. These discourse practices include:

- 1) Indicating one's own "partially knowing", "unknowing" or "intention to know" status to allow dialogic space for others to pitch in;
- 2) Addressing audience whether they are real or imaginary, such that others will be invited to a dialogic space;
- 3) Qualifying one's own stance to specify the conditions under which it is applicable to allow other conditions, i.e., a dialogic space for alternative stance;
- 4) Acknowledging the possibility of alternative voices and potential differences underlying them to allow a dialogic space of multiple voices;
- 5) Creating a common ground with meta-language to establish a dialogic space;
- 6) Relating intertextually to other sources or other users' stances to expand a dialogic space to other voices.

With these discourse practices, users can establish and expand dialogic space, and potentially achieve intersubjectivity. Additionally, the micro-analysis of threads provides an empirical description of discourse practices that make the processes of intersubjectivity possible during stance-taking for online deliberation, whether users converge or remain polarized in their views. These discourse practices include:

- 1) While asserting one's own subjectivity, indicating concession to others' subjectivity to show the possibility of integration of both subjectivities;
- 2) Using meta-language for clarification of one's own or others' subjectivity and epistemic status, and acknowledging the clarification;
- 3) Framing the on-going conversations as exploring differences and similarities;
- 4) Not fixating on the presence of evidence itself but the content within the evidence to unravel the possible common ground between different subjectivities;

In short, this thesis contributed to the existing knowledge of dialogic space and intersubjectivity with a shift to online discourse and with detailed linguistic evidence concluded from the analysis of online discussions.

10.5 Methodological contributions

The methodological contributions of this thesis lie in its innovation in two ways: (1) differentiating comments in online discussions to systematically examine online discourse; and (2) applying corpus linguistics with micro-analysis to investigate both general patterns and nuanced discourse practices in MOOC online discussions.

10.5.1 Dissecting online discussions

As far as I am aware, this is the first corpus study of online discourse and MOOC that differentiates the comments in online discussions at such a fine-grained level: new posts (initiating posts vs. independent posts), replies (first contributions vs. subsequent contributions). There have been studies beginning distinguishing posts and replies (Collins, 2019; Ksiazek & Lessard, 2016), and posts that receive replies vs. those that do not (Arguello et al., 2006; Burke et al., 2007; Crook et al., 2016; Rooderkerk & Pauwels, 2016) in order to examine users' interactions in online discussions. This large-scale corpus analysis of online discussions built on these previous studies to explore all the different types of comments at the same time. The resulting keywords and discourse practices drawn from the comparisons of initiating posts, independent posts and replies also justify this differentiation and further revealed the nuanced aspects of users' interactions in online discussions.

The identification of different kinds of replies – first contributions and subsequent contributions – which has not been done before, further informs the social dynamics within a thread and users' posting behaviours in online discussions. As can be seen in this thesis, most users seldom continue to engage in the same thread they have joined or initiated before. Yet, long threads are found more likely to be sustained by users' continued engagement in the same threads; that is

subsequent contributions. Furthermore, in general, first contributions are found to be expression of agreement whereas subsequent contributions are expressions of disagreement and words of responsiveness, suggesting that the two types of replies are indeed different in nature and justifying such a differentiation.

The differentiation of comments in online discussions also furthers our understanding of users' posting behaviours in online discussions, as shown by the categorization of users based on their contributions of different types of comments, extending previous characterization of users purely based on users' frequency of posting (Huang et al., 2014; Ruiz et al., 2011). This further confirms the validity of making such a fine-grained analysis of comments in online discussions, and future research may consider taking such differentiation of comments into account for their analysis of social interactions in online discussions.

10.5.2 Integrating keyword analysis and micro-analysis

Although corpus linguistics emphasizes both quantitative and qualitative analysis, qualitative analysis of concordance lines is at times limited in previous research, especially when analysing conversation threads in online discussions (e.g., Beers-Fägersten, 2008; Drasovean & Tagg, 2015). To overcome this limitation, this thesis successfully integrated keyword analysis and micro-analysis adapted from Conversation Analysis to examine users' replies to each other within a conversation thread. This way, not only can the discourse function of a reply keyword be better understood through in-depth analysis, but the micro-analysis is also driven by statistically significant keywords. For example, users' use of the reply keyword *but* for reassertion after concession with others' point of view, and its impact for negotiation was revealed by the micro-analysis of the thread in which two users engaging in a sustained conversation. Meanwhile, the micro-analysis of threads was informed by focusing on the reply keywords used for meta-language, and the analysis showed how development of threads can reach reconciliation or stalemate. Therefore, integration of keyword analysis and micro-analysis can be considered as one way of using a corpus approach to assist

discourse analysis in future research on conversation threads in online spaces (Partington, Duguid, & Taylor, 2013).

10.5.3 Corpus linguistics for analysing online discussions in MOOCs

This thesis illustrated that a corpus linguistic approach is useful for examining big language data available from online discussions such as MOOCs while enabling in-depth discourse analysis of actual language usage at the same time. This mixed methodology reveals both the general patterns and specific discourse practices in the online discussion. Furthermore, technically, as documented in Chapter 4, corpus tools encode meta-data and the hierarchical structures of online discussions such that each comment is still stored within their context and co-text, annotated with variables of interest. This corpus management is useful for a full integration and iteration of quantitative and qualitative analysis of the textual data of online discussions (Wegerif & Mercer, 1997). As has been shown in the analysis of this thesis, the quantitative component of keyword analysis or collocation analysis points to a specific keyword or collocate which discourse functions are to be examined qualitatively by concordancing. Given that the textual data is stored in their complete co-text, the concordancing allows expansion of co-text to the whole threads to facilitate qualitative discourse analysis, as have been conducted for micro-analysis in this thesis.

On one hand, this integration of quantitative and qualitative analysis can prevent the researchers' presumptions regarding the discourse functions of certain linguistic features, as in the case of LIWC (Pennebaker et al., 2001) used by Arguello et al. (2006) and Crook et al. (2016). For example, as shown in this thesis, the initiating keyword *wrong* does not necessarily carry a negative meaning but is used typically by users in the online discussion in FutureLearn as a way to express epistemic uncertainty, whereas the reply keyword *link* is not used to establish logical reasoning, but to refer to the URL posted.

On the other hand, if a certain pattern is spotted during the discourse analysis, it can in turn be subjected to quantitative analysis or concordancing in the corpus tools to establish if it is a

recurrent pattern, thus allowing an iterative process of both quantitative and qualitative analysis. For example, the phrase *agree to disagree* first came to my attention during the qualitative analysis of the reply keyword *agree*, then its importance in online discussion space was established by comparing its frequency to spoken and written word corpus, and micro analysis of threads. All these analyses attest to the methodological advantage of corpus linguistics that allows integration and iteration of quantitative and qualitative analysis, which could be deemed as a fully mixed methodology.

In short, this study built on previous corpus studies on online spaces and it applied established corpus methods to the field of MOOC research which thus far has not investigated users' textual contributions from the perspective of language use in context but reducing the language data to codes to infer users' thinking (Almatrafi & Johri, 2019; Wise et al., 2016). The data-driven approach also allows exploration of users' textual contributions in MOOC online discussions without any a priori framework, given the observation in this thesis suggest that it seems to be a mixed-genre, rather than purely educational. This study also expands on Collins' (2019) preliminary corpus analysis of one MOOC to a larger corpus consisting of 12 MOOCs and more in-depth analysis of discourse practices of different types of comments and micro-analysis of user-user interactions.

10.6 Practical Implications

The findings of this thesis reveal how users employ different discourse practices in their interaction in the online discussions through the mediation of internet and platform design. These findings have significant practical implications for users' agency in online spaces, as well as the possibility of utilizing FutureLearn MOOC discussions as a third space.

10.6.1 Discourse practices as one aspect of digital literacies

The findings in this thesis illustrate that, although constrained by the technological design of online spaces, users execute their agency via language to engage in information exchange, stance-taking

and socialization with others, or express themselves as engaging in user-content interactions or information distributors for others. This attests to the argument that users' agency in online spaces is equally, if not more, important than the technological design. The observation that some discourse practices may hinder their interactions with others, either in the case of not getting reply when they need it, or not engaging in constructive deliberation when they have opposing stances, further points to the importance of raising users' awareness regarding their discourse practices online. Therefore, language use should remain as one aspect of digital literacies that is worth cultivating in this digital era (Herring, 2004; Jones & Hafner, 2012; Tagg & Seargeant, 2019; Thorne, 2013).

The findings in this thesis could effectively inform users of their language practices and URL-posting in online discussions. To establish dialogic conversations with others, users can employ not only the linguistic features found, but also words other than the keywords, as long as they engage in the discourse practices useful for expanding dialogic spaces and facilitating intersubjectivity. These insights are afforded through the in-depth analysis of discourse directed by the two theoretical concepts, which are not available from previous studies that only stop at linguistic features or codes such as question, disagreement or on-topic (Arguello et al., 2006; Cui et al., 2017). Essentially, the thesis shows *how* to write a comment, not just *what* to write.

Admittedly, the users may not be readily aware of the theoretical explanation of discourse practices, or the use of *if*, modals, meta-language and other linguistic accounts. However, the discourse practices found useful for expanding dialogic space and facilitating intersubjectivity can be translated to different terminologies to inform users. Example suggestions include:

- 1) Try to recognize others' viewpoints and acknowledge what others have written;
- 2) Don't be too assertive or endeavor to defeat others;
- 3) Don't sound as if others are less knowledgeable than you;
- 4) Don't make sweeping generalizations;
- 5) Don't be afraid to express that you do not know or understand;

- 6) Don't fixate or repeat one single issue;
- 7) Don't repeatedly criticize others' ways of posting;
- 8) Use oral language to address others;
- 9) Write about the relevance of the URLs to the current discussions;
- 10) Comment on the content of the sources linked to the URLs posted, but not the URLs themselves;
- 11) Acknowledge others' personal situated experience.

The findings in this thesis not only provide insights for internet users but also managers or members of virtual communities, including those in Facebook, LinkedIn or WhatsApp groups, and other general groups for social support or shared interest, or groups hosted by private companies for customers' engagement. For example, the findings regarding discourse practices for initiating conversations and URL- posting might be particularly useful for those who would like to become a "Conversation Starter", "Conversation Booster", "Link Curator" in Facebook groups ("Facebook Group Badges | Facebook Community," n.d.). Members of other groups can also employ the discourse practices found in this thesis to keep conversations alive in their groups, while managers of an online group can identify problems in their groups and facilitate conflict based on these practices.

However, it should be noted that hard and fast rules for interactions might not be feasible in online discussions. As revealed in the micro-analysis, the communicative norms are co-constructed by users in their interactions and vary depending on the members of a community (Kleinke & Bos, 2015; Marra, 2012; Netz, 2014; Stommel & Koole, 2010; Tanskanen, 2007). This suggests that users might have to acquire different communicative norms whenever they join an online discussion. Sometimes, users even engage in metapragmatic discussions, i.e., discussions about how to discuss (Tanskanen, 2007), as shown in the cases of when disagreement should stop, if an URL is compulsory

evidence for stance-taking, or if there must be a “winner” in a disagreement. Although these metapragmatic discussions evolve from users’ disagreement over other issues, their presence suggests that it is important for users to co-construct the communicative norms in the online discussions and also to be aware of the underlying differences among them in terms of engagement in online discussions.

10.6.2 FutureLearn online discussions as a third space

The presence of disagreement not only highlights its importance for sustained interactions, but also the potential of FutureLearn online discussions as a dialogic space for multiple voices and intersubjectivity. Contentious issues have been raised in a few threads presented in this thesis, including childbirth benefits in finance-1, anthropogenic effect of climate change in moon-1, cause of Irish famine in soil-1, homeopathy in ancient-1. Most of these discussions evolve from user-user interactions on course-related topics, rather than being directly prompted by the discussion prompts or course contents. This suggests that FutureLearn online discussions could potentially be a third space (Wright, 2012), where users with different views are brought together incidentally because of MOOCs, rather than self-selecting into a homogenous group that share the same stance as in the case of echo chambers (Walter et al., 2018).

According to Wright (2012) and T. Graham (2010), it is common that online deliberation incidentally arises from everyday conversations, rather than a purposefully designed space driven by homogenous groups who already share the same opinions. Therefore, the fact that large number of users who might hold different views happen to be brought together due to a MOOC may give rise to such online deliberation. Besides, another significant feature of FutureLearn “discussion in context” allows users free to post in the margin of steps, thus promoting incidental occurrence of online deliberation. The incidental rise of new topics may go against the view of some educational researchers who have voiced their concerns over irrelevant topics in online discussions (Wise et al., 2016). However, Herring (1999) and Benwell & Stokoe (2006), as well as educational researchers

Wegerif (2010) and Faraj et al. (2011), argue that the multiplicity arising from the online discussions could actually be appealing to users. Furthermore, other researchers (Mercer, 2004; Potter, 2008; Ugoretz, 2005; Wright, 2012) contend that a digression can be an exploration that leads to productive discussions of different perspectives. Nonetheless, the success of an online discussion developing into a third space that facilitates online deliberation and intersubjectivity is largely shaped by users' discourse, as shown in this thesis.

Additionally, with users holding polarized stances engaging in online discussions, analysis of their discourse may help to reveal what underlies their strong stances, thus informing how to possibly suggest change on certain socially constructed perceptions on those issues (Jacobson et al., 2016; Koschack et al., 2015). For example, in this thesis, it is found that differing evidencing practices, such as presence of a URL mentioning authority vs. a real-life personal experience, may explain why opposing parties stick to their stance rather than acknowledging the possibility of other voices or attempting to integrate each other's valid points. Therefore, FutureLearn online discussions may potentially be utilized as a third space for understanding alternative voices on various divisive issue, such as anthropogenic climate change, racism, or vaccination. At the same time, analysis of users' discourse will reveal what underlies their persistent construal of the social world regarding these divisive issues, thus informing policy-making or educational program.

10.7 Limitations and future research

This large-scale corpus linguistic study integrated both quantitative and qualitative analysis to examine online discourse, revealing discourse practices of a dialogic nature in different types of comments and threads, as well as URL-posting in online discussions. Despite its multi-faceted investigation of the discourse practices in online spaces, limitations of this thesis are identified and future research are suggested.

10.7.1 Generalizability

As argued in Chapter 2, I chose the MOOC online discussion to investigate online discourse because of its growing popularity, interactive nature and the possibility that the discourse practices in the educational site will inform best practices for users' interactions online. Admittedly, this choice of research setting, regardless of how large the FL corpus is, render the findings rather limited such that they are not generalizable to other online spaces. Nonetheless, the findings remain statements about what can possibly be achieved by users' discourse practices, such that similar practices can be employed in other online spaces to facilitate dialogic conversations.

One way to explore the generalizability of the findings is to compare the FL corpus with other corpora of online discourse, such as Yahoo News Annotated Comments Corpus (Napoles et al., 2017) and e-language corpus (D. Knight et al., 2014). However, due to ethical concerns, the FL corpus compiled in this thesis cannot be made publicly available for other researchers to investigate or draw upon for comparison with other online spaces. Nonetheless, the findings in this thesis can still form the basis for future comparison. Future research can also further explore users' URL-postings and situations where they agree to disagree in other digital spaces to extend our understanding of online disagreement and polarization.

Another potential criticism towards the generalizability of current investigation of online discourse is the lack of information regarding users' background. However, as argued in Chapter 5, the current study did not aim to investigate online discourse of a certain group of users. Rather, it was to investigate online discourse as it occurs in an online space, which was successfully achieved by the data-driven approach taken in this thesis. Furthermore, as shown in the micro-analysis, users' characteristics and identities were rather fluid and best seen when enacted via discourse by users.

A final issue is the presence of super-posters. Similar to other online discussions (T. Graham & Wright, 2014; Poquet et al., 2018), some users were found to be more vocal and prolific in the FutureLearn MOOC online discussions. For example, one super-poster has been involved in quite a few threads presented in the micro-analysis. This user has been found voicing disagreement

relentlessly despite the appeal of other users to withdraw from the discussion in two threads, and has also been involved in numerous disagreements where they do not show acknowledgement of others' voices. Nonetheless, there are also times this user provides useful information in response to others' queries. Specifically, 98% of 368 contributions by this user are replies, suggesting the role of super-posters in responding to others, partially reducing the chance of a post not receiving a reply. Arguably the large number of contributions of super-posters may skew the quantitative analysis in this thesis. However, as shown in Chapter 5, the contributions of other users are large enough such that the corpus is representative of all users instead of over-representative of a few super-posters.

To some extent, the existence of super-posters represents what is happening in the online space, rather than a bias. If I have come across their comments very often in my analysis, not to mention other users who have been engaging in the MOOC discussion. Several studies have shown that they tend to dictate the topic and behave the same in other courses, despite the positive impact they may have exerted in the community (Huang et al., 2014; Lambiase, 2010). Future research can follow T. Graham and Wright (2014) to further analyse the discourse practices of different super-posters to understand their role in online discussions. Similar studies can also be conducted on selected individuals, such as one-time contributors and the other groups of users found in Chapter 5, to complement the current study that focuses on general patterns and discourse practices across all users.

10.7.2 Discourse as observed

As discussed in Chapter 4, corpus linguistic approach assumes that the recurrent patterns of discourse among language users (i.e., speakers, writers, online users) reflect their interaction and construal of the social world, although they themselves and their interlocutors may not necessarily be aware of it (McEnery & Hardie, 2012). At the same time, the micro-analysis adapts the principle of Conversation Analysis (Heritage, 2004), thereby, the interpretation of users' comments also takes into account the response they receive and the comments they respond to. To some extent,

language users might not be totally aware of what they are doing, and it is the researchers' task to identify these social practices based on recurrent patterns observable in users' language choices and their distinctive function in conversations.

Admittedly, despite this assumption, the thesis could further triangulate the observation and interpretation of discourse by interviewing the users' involved (Wagner & Herbel-Eisenmann, 2008). As argued in the analysis of disagreement in Chapter 8, it is possible that some users do not acknowledge the existence of alternative voices because they argue for the purposes of "winning" (Berryman-Fink, 1998; Felton et al., 2015; S. L. Graham, 2007). Furthermore, as concluded in Chapter 7, most of the discourse in the independent posts seem to reflect users' interactions with course content and are in response to discussion prompts. It has also been shown that users may construe the communicative norms differently in the online discussion, such as how to disagree, how to comment and provision of evidence. Admittedly, from discourse, we would not be able to deduce users' internal mental states. Therefore, an interview that elicits such information might be useful for examining this possibility, although interview data is similarly subjective. Users might not be aware of their intention and how they feel, forget about the situations when they employ certain discourse practices, or not be willing to reveal their intention. Any difference between what I observed and what users think will also warrant future research.

Lastly, to more reliably claim what discourse practices are more likely to render a post to receive a reply, a comparison between independent posts and initiating posts with similar content but different discourse practices is needed. However, it might not be feasible to sieve through large number of data and find such a match. Future research could follow an experimental paradigm by manipulating the comments users see and comparing how they will respond, as has been done by Concannon et al.(2017), Joyce et al. (2007), Pöldvere et al. (2016). However, it will not be possible in a natural setting without intruding into the users' interactions, and only relatively few discourse practices can be tested in such a controlled environment.

10.7.3 Other discourse practices for dialogic conversations

In this thesis, the threads under micro-analysis are mostly between two users who have opposing stances. This is partly due to the fact that with the two users coming back, my interpretation of each of their comments can be supported by their other comments in the thread, and it is more straightforward to identify who they are talking to compared to in a polylogue. Admittedly, there are possibly other discourse practices that can shape dialogic conversations in threads which are more polylogal than those I have analysed. For example, one long thread with 42 replies in ancient-1 consisted of multiple users sharing their own experience of sex education, although towards the end it becomes an exercise in stance-taking among a few users regarding what should be included in sex education. This suggests the potential role of experience talk in online conversations when it is started by users themselves, as has been investigated by Kääntä & Lehtinen (2016) and Jaworska (2018). In another long thread with 17 replies in nutrition-4, a stepmother sought advice and support from others for cultivating nutritious eating with her step-children, suggesting the potential role of advice-giving with hedging in online conversations. This thread is mainly her and a few users continuously engaging in turn-taking, rather than a dialogue between two users. All these are based on my preliminary reading and require further research to examine if there are other discourse practices which will be useful for expanding dialogic space and facilitating the process of intersubjectivity, especially in polylogal conversations. However, it should be acknowledged that long threads with multiple conversations interleaving can contain multiple discourse practices and can evolve from one topic to another, therefore it might be challenging to draw a conclusion.

10.8 Concluding Remarks

Technology has evolved over the years to advance our communications with each other, especially allowing us to engage in interactions with people we do not know in online spaces. However, it is important not to overlook the fundamental meaning-making resource – language – while we interact with others in online spaces. Thus, digital literacies should still include users' language

practices as one aspect. As has been argued throughout this thesis, users' discourse practices matter in initiating and sustaining dialogic conversations with others, as well as turning a disputation into a constructive exploration of differences. This is the agency of users in the online world even though it may be mediated by technological design. This research is timely given that nowadays we are offered platforms such as online spaces to express our views, yet we seem not to be engaging in dialogic conversations or deliberations with others who might have opposing views from us on various issues, ranging from climate change to dealing with Covid-19. Although it is hard to establish conversations with people we disagree, recognizing our differences and being mindful with one's words may be the start.

References

- Alario-Hoyos, C., Muñoz-Merino, P. J., Pérez-Sanagustín, M., Delgado Kloos, C., & Parada G., H. A. (2016). Who are the top contributors in a MOOC? Relating participants' performance and contributions. *Journal of Computer Assisted Learning*, 32(3), 232–243. <https://doi.org/10.1111/jcal.12127>
- Almatrafi, O., & Johri, A. (2019). Systematic review of discussion forums in Massive Open Online Courses (MOOCs). *IEEE Transactions on Learning Technologies*, 12(3), 413–428. <https://doi.org/10.1109/TLT.2018.2859304>
- Angouri, J., & Tseliga, T. (2010). “you HAVE NO IDEA WHAT YOU ARE TALKING ABOUT!” From e-disagreement to e-impoliteness in two online fora. *Journal of Politeness Research*, 6(1), 57–82. <https://doi.org/10.1515/JPLR.2010.004>
- Antaki, C., & Ardévol, E. (2005). “For she who knows who she is:” Managing accountability in online forum messages. *Journal of Computer-Mediated Communication*, 11, 114–132. <https://doi.org/10.1111/j.1083-6101.2006.00006.x>
- Anthony, L. (2017). AntConc. Tokyo, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net/software/antconc/>
- Arguello, J., Butler, B. S., Joyce, L., Kraut, R. E., Ling, K. S., & Wang, X. (2006). Talk to me: Foundations for Successful Individual-Group Interactions in Online Communities. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '06*, 959–968. <https://doi.org/10.1145/1124772.1124916>
- Baker, P. (2004). Querying keywords: questions of difference, frequency, and sense in keywords analysis. *Journal of English Linguistics*, 32(4), 346–359. <https://doi.org/10.1177/0075424204269894>
- Baker, P., & Egbert, J. (2018). Introduction. In P. Baker & J. Egbert (Eds.), *Triangulating Methodological Approaches in Corpus Linguistic Research* (1st ed.). Routledge.
- Baker, P., Gabrielatos, C., Khosravimik, M., Krzyżanowski, M., McEnery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*, 19(3), 273–306. <https://doi.org/10.1177/0957926508088962>
- Bakhtin, M. M. (1981). *The Dialogic Imagination: Four Essays by M. M. Bakhtin*. Austin, TX: University of Texas Press. <https://doi.org/10.2307/2068977>
- Baroni, M., & Evert, S. (2016). Statistics for Linguists with R-a SIGIL course Unit 2: Corpus Frequency Data & Statistical Inference. Retrieved December 8, 2020, from http://www.stefan-evert.de/SIGIL/sigil_R/materials/02_corpus_frequency_data.handout.pdf
- Baumgarten, N., & House, J. (2010). I think and I don't know in English as lingua franca and native English discourse. *Journal of Pragmatics*, 42(5), 1184–1200. <https://doi.org/10.1016/j.pragma.2009.09.018>
- Baym, N. K. (1996). Agreements and disagreements in a computer-mediated discussion. *Research on Language and Social Interaction*, 29(4), 315–345. <https://doi.org/10.1207/s15327973rlsi2904>

- Beers-Fägersten, K. (2008). A corpus approach to discursive constructions of a hip-hop identity. In A. Ädel & R. Reppen (Eds.), *Corpora and Discourse: The challenges of different settings* (pp. 211–240). John Benjamins Publishing Company.
- Bellander, T., & Landqvist, M. (2020). Becoming the expert constructing health knowledge in epistemic communities online. *Information Communication and Society*, 23(4), 507–522. <https://doi.org/10.1080/1369118X.2018.1518474>
- Benwell, B., & Stokoe, E. (2006). *Virtual Identities*. Edinburgh University Press. <https://doi.org/10.1002/9781444397222.ch8>
- Benwell, B., & Stokoe, E. H. (2002). Constructing discussion tasks in university tutorials: shifting dynamics and identities. *Discourse Studies*, 4(4), 429–453. <https://doi.org/10.1177/1461445602004004023>
- Berryman-Fink, C. (1998). Can we agree to disagree? Faculty-faculty Conflict. In S. Holton (Ed.), *Mending the Cracks in the Ivory Tower: Strategies for Conflict Management in Higher Education* (pp. 141–163). Bolton, MA: Anker Publishing Company.
- Beth, A. D., Jordan, M. E., Schallert, D. L., Reed, J. L. H., & Kim, M. (2015). Responsibility and generativity in online learning communities. *Interactive Learning Environments*, 23(4), 471–484. <https://doi.org/10.1080/10494820.2013.788035>
- Biber, D. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes*, 5(2), 97–116. <https://doi.org/10.1016/j.jeap.2006.05.001>
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371–405. <https://doi.org/10.1093/applin/25.3.371>
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow, United Kingdom: Pearson Education Limited.
- Bolander, B. (2012). Disagreements and agreements in personal/diary blogs: A closer look at responsiveness. *Journal of Pragmatics*, 44(12), 1607–1622. <https://doi.org/10.1016/j.pragma.2012.03.008>
- Bolander, B. (2013). *Language and Power in Blogs: Interaction, disagreements and agreements*. John Benjamins Publishing Company.
- Bou-Franch, P. (2012). Social interaction in YouTube text-Based polylogues : A Study of coherence. *Journal of Computer-Mediated Communication*, 17, 501–521. <https://doi.org/10.1111/j.1083-6101.2012.01579.x>
- Bou-Franch, P., & Garcés-Conejos Blitvich, P. (2014). Conflict management in massive polylogues: A case study from YouTube. *Journal of Pragmatics*, 73, 19–36. <https://doi.org/10.1016/j.pragma.2014.05.001>
- Brennan, S. E., & Ohaeri, J. O. (1999). Why do electronic conversations seem less polite? the costs and benefits of hedging. In *ACM SIGSOFT Software Engineering Notes* (Vol. 24, pp. 227–235). <https://doi.org/10.1145/295666.295942>
- British Psychological Society. (2017). *Ethics Guidelines for Internet-mediated Research*. British Psychological Society. <https://doi.org/INF206/1.2013>
- Brookes, G., & Baker, P. (2017). What does patient feedback reveal about the NHS? A mixed methods study of comments posted to the NHS Choices online service. *BMJ Open*, 7(4), 1–12. <https://doi.org/10.1136/bmjopen-2016-013821>

- Brown, P., & Levinson, S. C. (1987). *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- Bucholtz, M., & Hall, K. (2005). Identity and interaction: a sociocultural linguistic approach. *Discourse Studies*, 7(4–5), 585–614. <https://doi.org/10.1177/1461445605054407>
- Burke, M., & Kraut, R. (2008). Mind your P's and Q's: When politeness helps and hurts in online communities. *28th Annual CHI Conference on Human Factors in Computing Systems*, 3195–3200. <https://doi.org/10.1145/1358628.1358830>
- Burke, Moira, Joyce, E., Kim, T., Anand, V., & Kraut, R. (2007). Introductions and requests: Rhetorical strategies that elicit response in online communities. In *Proceedings of the 3rd Communities and Technologies Conference, C and T 2007* (pp. 21–39). https://doi.org/10.1007/978-1-84628-905-7_2
- Burnett, G., & Jaeger, P. T. (2008). Small worlds, lifeworlds, and information: The ramifications of the information behaviours of social groups in public policy and the public sphere. *Information Research*, 13(2), 1–18.
- Camiciottoli, B. C. (2009). “Just wondering if you could comment on that”: Indirect requests for information in corporate earnings calls. *Text and Talk*, 29(6), 661–681. <https://doi.org/10.1515/TEXT.2009.034>
- Cao, C., Caverlee, J., Lee, K., Ge, H., & Chung, J. (2015). Organic or organized? Exploring URL sharing behavior. *International Conference on Information and Knowledge Management, Proceedings, 19-23-Oct-*, 513–522. <https://doi.org/10.1145/2806416.2806572>
- Carter-Thomas, S., & Rowley-Jolivet, E. (2008). If conditionals in medical discourse: From theory to disciplinary practice. *Journal of English for Academic Purposes*, 7, 191–205.
- Carter, R., & McCarthy, M. (2006). *Cambridge Grammar of English: a comprehensive guide - spoken and written English Grammar and usage*. Cambridge: Cambridge University Press.
- Castelo, L. M. P., & Monaco, L. M. (2013). Conditionals and their functions in women's scientific writing. *Procedia - Social and Behavioral Sciences*, 95, 160–169. <https://doi.org/10.1016/j.sbspro.2013.10.635>
- Cavanagh, A. (2007). *Sociology in the Age of the Internet*. Maidenhead: Open University Press.
- Chen, G., & Chiu, M. M. (2008). Online discussion processes: Effects of earlier messages' evaluations, knowledge content, social cues and personal information on later messages. *Computers and Education*, 50(3), 678–692. <https://doi.org/10.1016/j.compedu.2006.07.007>
- Chen, G., Lo, C. K., & Hu, L. (2020). Sustaining online academic discussions: Identifying the characteristics of messages that receive responses. *Computers and Education*, 156(October 2019). <https://doi.org/10.1016/j.compedu.2020.103938>
- Cheng, W. (2004). Did you took from the minibar? What is the practical relevance of a corpus driven language study to practitioners in Hong-Kong's hotel industry? *Discourse in the Professions*, 141–166.
- Chiu, M. M. (2008). Flowing toward correct contributions during group problem solving: A statistical discourse analysis. *Journal of the Learning Sciences*, 17(3), 415–463. <https://doi.org/10.1080/10508400802224830>
- Coetzee, D., Fox, A., Hearst, M. A., & Hartmann, B. (2014). Should your MOOC forum use a reputation system? *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '14*, 1176–1187.

<https://doi.org/10.1145/2531602.2531657>

- Colaric, S., & Jonassen, D. (2001). Information equals knowledge, searching equals learning, and hyperlinking is good instruction. *Computers in the Schools*, 17(3–4), 33–46.
<https://doi.org/10.1300/J025v17n03>
- Collins Dictionary. (n.d.). Agree to disagree to agree to differ definition and meaning | Collins English Dictionary. Retrieved September 27, 2020, from
<https://www.collinsdictionary.com/dictionary/english/agree-to-disagree>
- Collins, L. (2019). *Corpus Linguistics for Online Communication: A Guide for Research*. New York: Routledge.
- Collins, L., & Nerlich, B. (2015). Examining user comments for deliberative democracy: A corpus-driven analysis of the climate change debate online. *Environmental Communication*, 9(2), 189–207. <https://doi.org/10.1080/17524032.2014.981560>
- Concannon, S., & Healey, P. G. T. (2015). Shifting Opinions: An experiment on agreement and disagreement in dialogue. In *19th SemDial Workshop on the Semantics and Pragmatics of Dialogue (goDIAL)*.
- Concannon, S., Healey, P. G. T., & Purver, M. (2017). Opening up and closing down discussion : Experimenting with epistemic status in conversation. In *39th annual Meeting of the Cognitive Science Society (CogSci 2017)* (pp. 1818–1823). Cognitive Science Society.
- Connor, L. G. O. (2013). Investors' information sharing and use in virtual communities. *Journal of the American Society for Information Science and Technology*, 64(1), 36–47.
<https://doi.org/10.1002/asi>
- Crook, B., Glowacki, E. M., Love, B., Jones, B. L., Macpherson, C. F., & Johnson, R. H. (2016). Hanging by a thread: exploring the features of nonresponse in an online young adult cancer survivorship support community. *Journal of Cancer Survivorship*, 10(1), 185–193.
<https://doi.org/10.1007/s11764-015-0465-8>
- Cui, Y., Jin, W. Q., & Wise, A. F. (2017). Humans and machines together: Improving characterization of large scale online discussions through Dynamic Interrelated Post and Thread Categorization (DIPTiC). *L@S 2017- Proceedings of the 4th 2017 ACM Conference on Learning at Scale*, 217–219.
- Dahlberg, L. (2001). The internet and democratic discourse: Exploring the prospects of online deliberative forums extending the public sphere. *Information, Communication & Society*, 4(4), 615–633. <https://doi.org/10.1080/13691180110097030>
- Dancygier, B., & Sweetser, E. (2000). Constructions with if, since, and because: Causality, epistemic stance, and clause order. *Cause, Condition, Concession, Contrast: Cognitive and Discourse Perspectives*, (1), 111–142.
- Danet, B. (2013). Flaming and linguistic impoliteness on a listserv. In S. Herring, D. Stein, & T. Virtanen (Eds.), *Pragmatics of Computer-Mediated Communication* (pp. 639–664). De Gruyter.
- de Maeyer, J. (2013). Towards a hyperlinked society: A critical review of link studies. *New Media and Society*, 15(5), 737–751. <https://doi.org/10.1177/1461444812462851>
- Delahunty, J. (2018). Connecting to learn, learning to connect: Thinking together in asynchronous forum discussion. *Linguistics and Education*, 46, 12–22.
<https://doi.org/10.1016/j.linged.2018.05.003>
- Dennen, V. P., & Wieland, K. (2007). From interaction to intersubjectivity: Facilitating online group

- discourse processes. *Distance Education*, 28(3), 281–297.
<https://doi.org/10.1080/01587910701611328>
- Drasovean, A., & Tagg, C. (2015). Evaluative language and its solidarity-building role on TED. com: An appraisal and corpus analysis. *Language@Internet*, (12), 1.
- Dron, J. (2014). Innovation and change: Changing how we change. In O. Zawacki-richter & T. Anderson (Eds.), *Online distance education : towards a research agenda* (pp. 237–266). AU Press, Athabasca University. <https://doi.org/10.15215/aupress/9781927356623.01> Cover
- Du Bois, J. W. (2007). The stance triangle. In R. Englebretson (Ed.), *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction* (pp. 139–182). Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/pbns.164.07du>
- Du Bois, J. W., & Kärkkäinen, E. (2012). Taking a stance on emotion: Affect, sequence, and intersubjectivity in dialogic interaction. *Text and Talk*, 32(4), 433–451.
<https://doi.org/10.1515/text-2012-0021>
- Duranti, A. (2004). Agency in Language. In A. Duranti (Ed.), *A Companion to Linguistic Anthropology* (pp. 451–473). Malden, MA: Blackwell Publishers.
<https://doi.org/doi:10.1002/9780470996522.ch20>
- Edgerly, S., Thorson, K., Bighash, L., & Hannah, M. (2016). Posting about politics: Media as resources for political expression on Facebook. *Journal of Information Technology and Politics*, 13(2), 108–125. <https://doi.org/10.1080/19331681.2016.1160267>
- Epstein, D., Farina, C., & Heidt, J. (2014). The value of words: Narrative as evidence in policy making, 10(2), 243–258.
- European Commission. (2018). 2018 reform of EU data protection rules. Retrieved from https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-changes_en.pdf
- Evert, S. (2008). Corpora and collocations. In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics. An International Handbook* (pp. 1212–1248). Berlin: Mouton de Gruyter.
<https://doi.org/10.1515/9783110213881.2.1212>
- Evert, S., & Hardie, A. (2011). Twenty-first century Corpus Workbench: Updating a query architecture for the new millennium. *Proceedings of the Corpus Linguistics 2011*, 1–21. Retrieved from <http://eprints.lancs.ac.uk/id/eprint/62721>
- Eysenbach, G., & Till, J. E. (2001). Ethical issues in qualitative research on internet communities. *British Medical Journal*, 323, 1103–1105. <https://doi.org/10.1136/bmj.323.7321.1103>
- Fairclough, N. (2003). *Analysing Discourse: Textual Analysis for social research*. Taylor & Francis. Retrieved from <http://eprints.lancs.ac.uk/4020/>
- Faraj, S., Jarvenpaa, S. L., & Majchrzak, A. (2011). Knowledge collaboration in online communities. *Organization Science*, 22(5), 1224–1239. <https://doi.org/10.1287/orsc.1100.0614>
- Fayard, A. L., & DeSanctis, G. (2005). Evolution of an online forum for knowledge management professionals: A language game analysis. *Journal of Computer-Mediated Communication*, 10(4), 1–36. <https://doi.org/10.1111/j.1083-6101.2005.tb00265.x>
- Felton, M., Garcia-Mila, M., Villarroel, C., & Gilabert, S. (2015). Arguing collaboratively: Argumentative discourse types and their potential for knowledge building. *British Journal of Educational Psychology*, (October 2016), 372–386. <https://doi.org/10.1111/bjep.12078>

- Ferguson, G. (2001). If you pop over there: A corpus-based study of conditionals in medical discourse. *English for Specific Purposes*, 20(1), 61–82. [https://doi.org/10.1016/S0889-4906\(99\)00027-7](https://doi.org/10.1016/S0889-4906(99)00027-7)
- Ferguson, R., & Clow, D. (2015). Consistent commitment: Patterns of engagement across time in Massive Open Online Courses (MOOCs). *Journal of Learning Analytics*, 2(3), 55–80. <https://doi.org/10.18608/jla.2015.23.5>
- Ferguson, R., & Sharples, M. (2014). Innovative pedagogy at massive scale: Teaching and learning in MOOCs. In C. Rensing, S. de Freitas, T. Ley, & P. J. M.- Merino (Eds.), *Open Learning and Teaching in Educational Communities, proceedings of 9th European Conference on Technology Enhanced Learning (EC-TEL 2014), Graz, Austria, September 16-19*. (pp. 98–111). Heidelberg: Springer.
- Flowerdew, L. (2008). *Corpus-based Analyses of the Problem–Solution Pattern*. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Freelon, D. (2015). Discourse architecture, ideology, and democratic norms in online political discussion. *New Media and Society*, 17(5), 772–791. <https://doi.org/10.1177/1461444813513259>
- Friess, D., & Eilders, C. (2015). A systematic review of online deliberation research. *Policy and Internet*, 7(3), 319–339. <https://doi.org/10.1002/poi3.95>
- Frobenius, M., & Harper, R. (2015). Tying in comment sections: The production of meaning and sense on Facebook: The self-explicative organization of communication acts on and through Facebook. *Semiotica*, 2015(204), 121–143. <https://doi.org/10.1515/sem-2014-0081>
- Fung, L., & Carter, R. (2007). Discourse markers and spoken english: Native and learner use in pedagogic settings. *Applied Linguistics*, 28(3), 410–439. <https://doi.org/10.1093/applin/amm030>
- FutureLearn. (2014). Terms and conditions - Research Ethics. Retrieved from <https://www.futurelearn.com/info/terms/research-ethics-for-futurelearn>
- FutureLearn. (2019). Terms and conditions - Privacy. Retrieved from <https://www.futurelearn.com/info/terms/privacy-policy>
- Gabrielatos, C. (2018). Keyness analysis: Nature, metrics and techniques. In C. Taylor & A. Marchi (Eds.), *Corpus Approaches to Discourse: A Critical Review* (pp. 225–258). Routledge. <https://doi.org/10.4324/9781315179346>
- Gabrielatos, C., & Baker, P. (2008). Fleeing, sneaking, flooding: A corpus analysis of discursive constructions of refugees and Asylum Seekers in the UK Press, 1996–2005. *Journal of English Linguistics*, 36(1), 5–38. <https://doi.org/10.1177/0075424207311247>
- Gallagher, S. E., & Savage, T. (2016). Observing URL sharing behaviour in Massive Online Open Courses. *L@S 2016 - Proceedings of the 3rd 2016 ACM Conference on Learning at Scale*, (September), 129–132. <https://doi.org/10.1145/2876034.2893387>
- Gasevic, D., Kovanović, V., Joksimović, S., & Siemens, G. (2014). Where is research on Massive Open Online Courses headed? A data analysis of the MOOC Research Initiative. *The International Review of Research in Open and Distributed Learning*, 15, 134–176. Retrieved from <http://web.b.ebscohost.com.ezproxy.auckland.ac.nz/ehost/pdfviewer/pdfviewer?vid=20&sid=14aeff05-1676-41f2-bd73-7ed87fb64741%40sessionmgr112&hid=102>
- Gibson, W. (2009). Negotiating textual talk: conversation analysis, pedagogy and the organisation of

- online asynchronous discourse. *British Educational Research Journal*, 35(5), 705–721.
<https://doi.org/10.1080/01411920802688754>
- Giglietto, F., Righetti, N., Rossi, L., & Marino, G. (2020). It takes a village to manipulate the media: coordinated link sharing behavior during 2018 and 2019 Italian elections. *Information Communication and Society*, 23(6), 867–891. <https://doi.org/10.1080/1369118X.2020.1739732>
- Giles, D., Stommel, W., Paulus, T., Lester, J., & Reed, D. (2014). Microanalysis of online data: The methodological development of “digital CA.” *Discourse, Context and Media*, 7, 45–51.
<https://doi.org/10.1016/j.dcm.2014.12.002>
- Gillani, N., & Eynon, R. (2014). Communication patterns in massively open online courses. *Internet and Higher Education*, 23, 18–26. <https://doi.org/10.1016/j.iheduc.2014.05.004>
- Gilster, P. (1997). *Digital Literacy*. New York: Wiley Computer Publications.
- Grabill, J. T., & Pigg, S. (2012). Messy rhetoric: Identity performance as rhetorical agency in online public forums. *Rhetoric Society Quarterly*, 42(2), 99–119.
<https://doi.org/10.1080/02773945.2012.660369>
- Grabowski, Ł. (2015). Keywords and lexical bundles within english pharmaceutical discourse: A corpus-driven description. *English for Specific Purposes*, 38, 23–33.
<https://doi.org/10.1016/j.esp.2014.10.004>
- Graham, S. L. (2007). Disagreeing to agree: Conflict, (im)politeness and identity in a computer-mediated community. *Journal of Pragmatics*, 39(4), 742–759.
<https://doi.org/10.1016/j.pragma.2006.11.017>
- Graham, T. (2010). The Use of Expressives in Online Political Talk: Impeding or Facilitating the Normative Goals of Deliberation?, 26–41.
- Graham, T., Jackson, D., & Wright, S. (2016). ‘We need to get together and make ourselves heard’: everyday online spaces as incubators of political action. *Information Communication and Society*, 19(10), 1373–1389. <https://doi.org/10.1080/1369118X.2015.1094113>
- Graham, T., & Wright, S. (2014). Discursive equality and everyday talk online: The impact of “superparticipants.” *Journal of Computer-Mediated Communication*, 19, 625–642.
<https://doi.org/10.1111/jcc4.12016>
- Grant, L. E. (2010). A corpus comparison of the use of I don’t know by British and New Zealand speakers. *Journal of Pragmatics*, 42(8), 2282–2296.
<https://doi.org/10.1016/j.pragma.2010.01.004>
- Gries, S. T. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, 13(4), 403–437. <https://doi.org/10.1075/ijcl.13.4.02gri>
- Guiller, J., & Durndell, A. (2006). “I totally agree with you”: Gender interactions in educational online discussion groups. *Journal of Computer Assisted Learning*, 22(5), 368–381.
<https://doi.org/10.1111/j.1365-2729.2006.00184.x>
- Hall, B. (2010). Interaction is insufficient: Why we need intersubjectivity in course room discourse. *Journal of ELearning and Online Teaching*, 1(12), 1–16.
- Hardie, A. (2012). CQPweb — combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, 17(3), 380–409.
<https://doi.org/10.1075/ijcl.17.3.04har>
- Heritage, J. (2004). Conversation analysis: Practices and methods. In D. Silverman (Ed.), *Qualitative*

- Research: Theory, Method and Practice* (pp. 222–246). Sage Publications.
- Heritage, J. (2012). Epistemics in action: Action formation and territories of knowledge. *Research on Language & Social Interaction*, 45(1), 1–29. <https://doi.org/10.1080/08351813.2012.646685>
- Herring, S. (1999). Interactional coherence in CMC. *Journal of Computer-Mediated Communication*, 4(4), JCMC444. <https://doi.org/10.1111/j.1083-6101.1999.tb00106.x>
- Herring, S. (2001). Computer-mediated discourse. In D. Schiffrin, D. Tannen, & H. Hamilton (Eds.), *The Handbook of Discourse Analysis* (pp. 612–634). Oxford: Blackwell Publishers. <https://doi.org/10.1017/CBO9780511805080.016>
- Herring, S. (2004). Computer-mediated discourse analysis: An approach to researching online communities. *Designing for Virtual Communities in the Service of Learning*, 338–376. <https://doi.org/10.1017/CBO9780511805080.016>
- Herring, S. (2013). Discourse in Web2.0: Familiar, reconfigured and emergent. In D. & Tannen & A. M. Tester (Eds.), *Discourse 2.0: Language and new media* (pp. 1–29). Washington, DC: Georgetown University Press. Retrieved from <http://info.ils.indiana.edu/~herring/GURT.2011.prepub.pdf>
- Hew, K. F. (2016). Promoting engagement in online courses: What strategies can we learn from three highly rated MOOCs. *British Journal of Educational Technology*, 47(2), 320–341. <https://doi.org/10.1111/bjet.12235>
- Hew, K. F., & Cheung, W. S. (2014). Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. *Educational Research Review*, 12, 45–58. <https://doi.org/10.1016/j.edurev.2014.05.001>
- Hewings, A., Coffin, C., & North, S. (2009). E- Conferencing: Corpus and Discourse Insights. In M. Charles, S. Hunston, & D. Pecorari (Eds.), *Academic Writing: At the Interface of Corpus and Discourse* (Vol. 53, pp. 1689–1699). London: Continuum. <https://doi.org/10.1017/CBO9781107415324.004>
- Hewitt, J. (2003). How habitual online practices affect the development of asynchronous discussion threads. *Journal of Educational Computing Research*, 28(1), 31–45. <https://doi.org/10.2190/PMG8-A05J-CUH1-DK14>
- Himelboim, I. (2008). Reply distribution in online discussions: A comparative network analysis of political and health newsgroups. *Journal of Computer-Mediated Communication*, 14(1), 156–177. <https://doi.org/10.1111/j.1083-6101.2008.01435.x>
- Himelboim, I., Gleave, E., & Smith, M. (2009). Discussion catalysts in online political discussions: Content importers and conversation starters. *Journal of Computer-Mediated Communication*, 14(4), 771–789. <https://doi.org/10.1111/j.1083-6101.2009.01470.x>
- Holmes, J., & Nesi, H. (2009). Verbal and mental processes in academic disciplines. *Academic Writing: At the Interface of Corpus and Discourse*, (August), 58–72.
- Horsella, M., & Sindermann, G. (1992). Aspects of scientific discourse: Conditional argumentation.
- Huang, J., Dasgupta, A., Ghosh, A., Manning, J., & Sanders, M. (2014). Superposter behavior in MOOC forums. *Proceedings of the First ACM Conference on Learning @ Scale Conference - L@S '14*, 117–126. <https://doi.org/10.1145/2556325.2566249>
- Hunt, D., & Harvey, K. (2015). Health communication and corpus linguistics: Using corpus tools to analyse eating disorder discourse online. In P. Baker & T. McEnery (Eds.), *Corpora and Discourse Studies. Palgrave Advances in Language and Linguistics* (pp. 134–146). London:

Palgrave Macmillan.

- Hyland, K. (2005). Stance and engagement: a model of interaction in academic discourse. *Discourse Studies*, 7(2), 173–192. <https://doi.org/10.1177/1461445605050365>
- Jacobson, S., Myung, E., & Johnson, S. L. (2016). Open media or echo chamber: the use of links in audience discussions on the Facebook Pages of partisan news organizations. *Information Communication and Society*, 19(7), 875–891. <https://doi.org/10.1080/1369118X.2015.1064461>
- Janier, M., & Reed, C. (2017). I didn't say that! Uses of SAY in mediation discourse. *Discourse Studies*, 19(6), 619–647. <https://doi.org/10.1177/1461445617715180>
- Jaworska, S. (2018). "Bad" mums tell the "untellable": Narrative practices and agency in online stories about postnatal depression on Mumsnet. *Discourse, Context & Media*, 25, 25–33. <https://doi.org/10.1016/j.dcm.2017.11.002>
- Jenkins, M., & Dragojevic, M. (2013). Explaining the process of resistance to persuasion: A politeness theory-based approach. *Communication Research*, 40(4), 559–590. <https://doi.org/10.1177/0093650211420136>
- Jeong, A., & Frazier, S. (2008). How day of posting affects level of critical discourse in asynchronous discussions and computer-supported collaborative argumentation. *British Journal of Educational Technology*, 39(5), 875–887. <https://doi.org/10.1111/j.1467-8535.2007.00789.x>
- Jones, R. H., & Hafner, C. A. (2012). *Understanding Digital Literacies: A Practical Introduction*. Taylor & Francis. <https://doi.org/10.1126/science.264.5166.1783-a>
- Jordan, M. P. (1998). The power of negation in English: Text, context and relevance. *Journal of Pragmatics*, 29(6), 705–752. [https://doi.org/10.1016/S0378-2166\(97\)00086-6](https://doi.org/10.1016/S0378-2166(97)00086-6)
- Joyce, E., & Kraut, R. E. (2006). Predicting continued participation in newsgroups. *Journal of Computer-Mediated Communication*, 11(3), 723–747. <https://doi.org/10.1111/j.1083-6101.2006.00033.x>
- Kääntä, L., & Lehtinen, E. (2016). Patterns of experience talk and argumentation in digital peer learning discussions. *Language@Internet*, 13, 1–15.
- Kärkkäinen, E. (2003). *Epistemic stance in English conversation: A description of its interactional functions, with a focus on I think*. John Benjamins Publishing Company.
- Kärkkäinen, E. (2006). Stance taking in conversation: From subjectivity to intersubjectivity. *Text and Talk*, 26(6), 699–731. <https://doi.org/10.1515/TEXT.2006.029>
- Kellogg, S. B., Booth, S., & Oliver, K. M. (2014). A social network perspective on peer support learning in MOOCs for educators. *International Review of Research in Open and Distance Learning*, 15(5), 263–289. <https://doi.org/10.19173/irrodl.v15i5.1852>
- Khan, M. L. (2017). Social media engagement: What motivates user participation and consumption on YouTube? *Computers in Human Behavior*, 66, 236–247. <https://doi.org/10.1016/j.chb.2016.09.024>
- Kiernan, P. (2018). Making the link: Intertextual resources in an online forum. In *Language, Identity and Cycling in the New Media Age* (pp. 149–164). <https://doi.org/10.1057/978-1-137-51951-1>
- Kilgarrieff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., ... Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1(1), 7–36. <https://doi.org/10.1007/s40607-014-0009-9>

- Kleinke, S. (2010). Interactive aspects of computer-mediated communication: “Disagreement” in an English and a German public news group. In S. Tanskanen (Ed.), *Discourses in Interaction* (pp. 195–223). John Benjamins Publishing Company.
- Kleinke, S., & Bos, B. (2015). Intergroup rudeness and the metapragmatics of its negotiation in online discussion fora. *Pragmatics*, 25(1), 47–71. <https://doi.org/10.1075/prag.25.1.03kle>
- Knight, D., Adolphs, S., & Carter, R. (2014). CANELC: Constructing an e-language Corpus. *Corpora*, 9, 29–56. <https://doi.org/10.3366/cor.2014.0050>
- Knight, S., & Littleton, K. (2016). Dialogue as Data in Learning Analytics for Productive Educational Dialogue. *Journal of Learning Analytics*, 2(3), 111–143. <https://doi.org/10.18608/jla.2015.23.7>
- Koschack, J., Weibezahl, L., Friede, T., Himmel, W., Makedonski, P., & Grabowski, J. (2015). Scientific versus experiential evidence: Discourse analysis of the chronic cerebrospinal venous insufficiency debate in a multiple sclerosis forum. *Journal of Medical Internet Research*, 17(7), e159. <https://doi.org/10.2196/jmir.4103>
- Koteyko, N., Jaspal, R., & Nerlich, B. (2013). Climate change and “climategate” in online reader comments: a mixed methods study. *The Geographical Journal*, 179(1), 74–86. <https://doi.org/10.1111/j.>
- Ksiazek, T. B., & Lessard, K. (2016). User engagement with online news: Conceptualizing interactivity and exploring the relationship between online news videos and user comments. *New Media and Society*, 18(3), 502–520. <https://doi.org/10.1177/1461444814545073>
- Laflen, A., & Fiorenza, B. (2012). “Okay, My Rant is Over”: The language of emotion in computer-mediated communication. *Computers and Composition*, 29(4), 296–308. <https://doi.org/10.1016/j.compcom.2012.09.005>
- Lambiase, J. J. (2010). Hanging by a thread: Topic development and death in an online discussion of breaking news. *Language@Internet*, 7, Article 9.
- Landqvist, M. (2016). Sense and sensibility – Online forums as epistemic arenas. *Discourse, Context and Media*, 13, 98–105. <https://doi.org/10.1016/j.dcm.2016.05.007>
- Langlotz, A., & Locher, M. A. (2012). Ways of communicating emotional stance in online disagreements. *Journal of Pragmatics*, 44(12), 1591–1606. <https://doi.org/10.1016/j.pragma.2012.04.002>
- Lankshear, C., & Knobel, M. (2006). Digital literacy and digital literacies. *Nordic Journal of Digital Literacy*, 1(April 2016), 12–24. <https://doi.org/10.1108/EL-05-2015-0076>
- Lapadat, J. (2002). Written interaction: A key component in online learning. *Journal of Computer-Mediated Communication*, 7(4). <https://doi.org/10.1111/j.1083-6101.2002.tb00158.x>
- Lapadat, J. (2007). Discourse devices used to establish community, increase coherence, and negotiate agreement in an online university course. *The Journal of Distance Education / Revue de l'Éducation à Distance*, 21(3), 59–92. <https://doi.org/Article>
- Laurillard, D. (2001). *Rethinking University Teaching: A Conversational Framework for the Effective Use of Learning Technologies*. Routledge.
- Laurillard, D. (2012). *Teaching as a design science: Building pedagogical patterns for learning and technology* (1st ed.). Routledge.
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word Frequencies in Written and Spoken English: Based on the British National Corpus*. London: Longman.

- Lewiński, M. (2010). Collective Argumentative Criticism in Informal Online Discussion Forums. *Argumentation and Advocacy*, 47(2), 86–105. <https://doi.org/10.1080/00028533.2010.11821740>
- Lewiński, M. (2013). Debating multiple positions in multi-party online deliberation: Sides, positions, and cases. *Journal of Argumentation in Context*, 2(1), 151–177. <https://doi.org/10.1075/jaic.2.1.07lew>
- Lewis, D. M. (2005). Arguing in English and French asynchronous online discussion. *Journal of Pragmatics*, 37(11 SPEC. ISS.), 1801–1818. <https://doi.org/10.1016/j.pragma.2005.02.014>
- Lijffijt, J., & Gries, S. T. (2012). Correction to Stefan Th. Gries' "Dispersions and adjusted frequencies in corpora", *International Journal of Corpus Linguistics*. *International Journal of Corpus Linguistics*, 17(1), 147–149. <https://doi.org/10.1075/ijcl.17.1.08lij>
- Lindström, J. K., & Londen, A. M. (2013). Concession and reassertion: On a dialogic discourse pattern in conversation. *Text and Talk*, 33(3), 331–352. <https://doi.org/10.1515/text-2013-0015>
- Littleton, K., & Whitelock, D. (2005). The negotiation and co-construction of meaning and understanding within a postgraduate online learning community. *Learning, Media and Technology*, 30(2), 147–164. <https://doi.org/10.1080/17439880500093612>
- Liu, P., & Liu, H. (2017). Creating common ground: The role of metapragmatic expressions in BELF meeting interactions. *Journal of Pragmatics*, 107, 1–15. <https://doi.org/10.1016/j.pragma.2016.10.006>
- Liyanagunawardena, T. R., & Williams, S. A. (2016). Elderly learners and Massive Open Online Courses: A review. *Interactive Journal of Medical Research*, 5(1), e1. <https://doi.org/10.2196/ijmr.4937>
- Loncar, M., Barrett, N. E., & Liu, G. (2014). Towards the refinement of forum and asynchronous online discussion in educational contexts worldwide : Trends and investigative approaches within a dominant research paradigm. *Computers & Education*, 73, 93–110. <https://doi.org/10.1016/j.compedu.2013.12.007>
- Lorenzo-Dus, N., Garcés-Conejos Blitvich, P., & Bou-Franch, P. (2011). On-line polylogues and impoliteness: The case of postings sent in response to the Obama Reggaeton YouTube video. *Journal of Pragmatics*, 43(10), 2578–2593. <https://doi.org/10.1016/j.pragma.2011.03.005>
- Louwerse, M. M., Crossley, S. A., & Jeuniaux, P. (2008). What if? Conditionals in educational registers. *Linguistics and Education*, 19(1), 56–69. <https://doi.org/10.1016/j.linged.2008.01.001>
- Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The Spoken BNC2014: designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics*, 22(3), 319–344.
- Malavasi, D., & Mazzi, D. (2010). History v. marketing: Keywords as a clue to disciplinary epistemology. *Keyness in Texts*, 169–184.
- Manipuspika, Y. S. (2019). Workplace Discourse: Constructing Social Identity of Lecturers Through WhatsApp Group. *International Journal of Language and Linguistics*, 7(1), 13. <https://doi.org/10.11648/j.ijll.20190701.13>
- Marcoccia, M. (2004). On-line polylogues: Conversation structure and participation framework in internet newsgroups. *Journal of Pragmatics*, 36(1), 115–145. [https://doi.org/10.1016/S0378-2166\(03\)00038-9](https://doi.org/10.1016/S0378-2166(03)00038-9)
- Marra, M. (2012). Disagreeing without being disagreeable: Negotiating workplace communities as an

- outsider. *Journal of Pragmatics*, 44(12), 1580–1590.
<https://doi.org/10.1016/j.pragma.2012.06.009>
- Martin, J. R., & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. Palgrave Macmillan.
- Mayes, J. T. (2015). Still to learn from vicarious learning. *E-Learning and Digital Media*, 12(3–4), 361–371. <https://doi.org/10.1177/2042753015571839>
- McDonald, D., & Woodward-Kron, R. (2016). Member roles and identities in online support groups: Perspectives from corpus and systemic functional linguistics. *Discourse and Communication*, 10(2), 157–175. <https://doi.org/10.1177/1750481315615985>
- McEnery, T. (2016). Keywords. In P. Baker & J. Egbert (Eds.), *Triangulating Methodological Approaches in Corpus Linguistic Research*. New York, New York, USA: Taylor & Francis.
- McEnery, T., & Hardie, A. (2012). Corpus linguistics: Method, theory and practice, 312.
<https://doi.org/10.1017/CBO9781107415324.004>
- Mercer, N. (2004). Sociocultural discourse analysis: analysing classroom talk as a social mode of thinking. *Journal of Applied Linguistics*, 1(2), 137–168.
<https://doi.org/10.1558/japl.2004.1.2.137>
- Meyer, S. B., Violette, R., Aggarwal, R., Simeoni, M., Macdougall, H., & Waite, N. (2019). Vaccine hesitancy and Web 2.0: Exploring how attitudes and beliefs about influenza vaccination are exchanged in online threaded user comments. *Vaccine*, 37(13), 1769–1774.
<https://doi.org/10.1016/j.vaccine.2019.02.028>
- Moore, J. (2013). The use of if/ then constructions in the achievement of information provision and client empowerment. *L'organizing : Une Question de Langage, Discours et Communication*.
- Myers, G. (2007). Enabling talk: How the facilitator shapes a focus group. *Text and Talk*, 27(1), 79–105. <https://doi.org/10.1515/TEXT.2007.004>
- Myers, G. (2009). *Discourse of Blogs and Wikis. Discourse of Blogs and Wikis*. Bloomsbury Publishing.
- Myers, G. (2010). Stance-taking and public discussion in blogs. *Critical Discourse Studies*, 7(4), 263–275. <https://doi.org/10.1080/17405904.2010.511832>
- Napoles, C., Tetreault, J., Rosato, E., Provenziale, B., & Pappu, A. (2017). Finding good conversations online: The Yahoo News Annotated Comments Corpus. *Proceedings of LAW 2017*, 13–23. Retrieved from <http://www.aclweb.org/anthology/W/W17/W17-08.pdf#page=25%0Ahttp://www.aclweb.org/anthology/W17-0802>
- Nathan, M. J., Eilam, B., & Kim, S. (2007). To disagree, we must also agree: How intersubjectivity structures and perpetuates discourse in a mathematics classroom. *The Journal of the Learning Sciences*, 16(4), 523–563.
- Netz, H. (2014). Disagreement patterns in gifted classes. *Journal of Pragmatics*, 61, 142–160.
<https://doi.org/10.1016/j.pragma.2013.09.007>
- O’Keeffe, A., & Walsh, S. (2012). Applying corpus linguistics and conversation analysis in the investigation of small group teaching in higher education. *Corpus Linguistics and Linguistic Theory*, 8(1), 159–181. https://doi.org/10.1007/978-3-319-33272-7_13
- Oeldorf-Hirsch, A., & Sundar, S. S. (2015). Posting, commenting, and tagging: Effects of sharing news stories on Facebook. *Computers in Human Behavior*, 44, 240–249.
<https://doi.org/10.1016/j.chb.2014.11.024>

- Oh, S., Oh, J. S., & Shah, C. (2008). The use of information sources by internet users in answering questions. In *Proceedings of the American Society for Information Science and Technology* (pp. 1–13).
- Pace, S., Buzzanca, S., & Fracocchi, L. (2016). The structure of conversations on social networks: Between dialogic and dialectic threads. *International Journal of Information Management*, 36(partA), 1144–1151. <https://doi.org/10.1016/j.ijinfomgt.2016.04.001>
- Paquot, M., & Bestgen, Y. (2009). Distinctive words in academic writing: A comparison of three statistical tests for keyword extraction. *Corpora: Pragmatics and Discourse. Papers from the 29th International Conference on English Language Research on Computerized Corpora (ICAME 29), Ascona, Switzerland, 14–18 May 2008*, 247–269. https://doi.org/10.1163/9789042029101_014
- Partington, A., Duguid, A., & Taylor, C. (2013). *Patterns and Meanings in Discourse: Theory and Practice in Corpus-assisted Discourse*. Amsterdam & Philadelphia, PA: John Benjamins Publishing Company.
- Paulus, T. M. (2006). Challenge or connect? Dialogue in online learning environments. *Journal of Computing in Higher Education*, 18(1), 3–29. <https://doi.org/10.1007/BF03032722>
- Pendry, L. F., & Salvatore, J. (2015). Individual and social benefits of online discussion forums. *Computers in Human Behavior*, 50, 211–220. <https://doi.org/10.1016/j.chb.2015.03.067>
- Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic Inquiry and Word Count (LIWC): LIWC2001 Manual*. Mahwah: Erlbaum Publishers.
- Pöldvere, N., Fuoli, M., & Paradis, C. (2016). A study of dialogic expansion and contraction in spoken discourse using corpus and experimental techniques. *Corpora*, 11(2), 191–225. <https://doi.org/10.3366/cor.2016.0092>
- Polletta, F., Chen, P. C. B., & Anderson, C. (2009). Is information good for democracy? Link-posting in an online forum. *Journal of Public Deliberation*, 5(1), Article 2.
- Pomerantz, A. (1984). Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. *Structures of Social Action*, 57–101. <https://doi.org/DOI:10.1017/CBO9780511665868.008>
- Poquet, O., Dowell, N., Brooks, C., & Dawson, S. (2018). Are MOOC forums changing? In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK '18)* (pp. 340–349). New York, New York, USA: ACM. <https://doi.org/https://doi.org/10.1145/3170358.3170416>
- Potter, A. (2008). Interactional coherence in asynchronous learning networks: A rhetorical approach. *Internet and Higher Education*, 11(2), 87–97. <https://doi.org/10.1016/j.iheduc.2008.05.001>
- Potts, A., & Semino, E. (2017). Healthcare professionals' online use of violence metaphors for care at the end of life in the US: A corpus-based comparison with the UK. *Corpora*, 12(1), 55–84. <https://doi.org/10.3366/cor.2017.0109>
- Potts, A., Simm, W., Whittle, J., & Unger, J. W. (2014). Exploring “success” in digitally augmented activism: A triangulated approach to analyzing UK activist Twitter use. *Discourse, Context and Media*, 6, 65–76. <https://doi.org/10.1016/j.dcm.2014.08.008>
- Preece, J., & Maloney-Krichmar, D. (2005). Online communities: Design, theory, and practice. *Journal of Computer-Mediated Communication*, 10(4), 1–10.
- Publication Manual of the American Psychological Association*. (2020) (7th ed.). Washington, DC:

- American Psychological Association. 2020. Retrieved from <https://apastyle.apa.org/products/publication-manual-7th-edition>
- R core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org/>
- Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, 1(4), 519–549.
- Rees-Miller, J. (2000). Power, severity, and context in disagreement. *Journal of Pragmatics*, 32, 1087–1111.
- Riordan, E., & Murray, L. (2010). A corpus-based analysis of online synchronous and asynchronous modes of communication within language teacher education. *Classroom Discourse*, 1(2), 181–198. <https://doi.org/10.1080/19463014.2010.514728>
- Rizvi, S., Rienties, B., & Khoja, S. A. (2019). The role of demographics in online learning; A decision tree based approach. *Computers and Education*, 137(January), 32–47. <https://doi.org/10.1016/j.compedu.2019.04.001>
- Robinson, J. D. (2009). Managing counterinforming : An interactional practice for soliciting information that facilitates reconciliation of speakers' incompatible positions. *Human Communication Research*, 35, 561–587. <https://doi.org/10.1111/j.1468-2958.2009.01363.x>
- Rooderkerk, R. P., & Pauwels, K. H. (2016). No comment?! The drivers of reactions to online posts in professional groups. *Journal of Interactive Marketing*, 35, 1–15. <https://doi.org/10.1016/j.intmar.2015.12.003>
- Rourke, L., & Kanuka, H. (2007). Barriers to online critical discourse. *International Journal of Computer-Supported Collaborative Learning*, 2(1), 105–126. <https://doi.org/10.1007/s11412-007-9007-3>
- Ruiz, C., Domingo, D., Micó, J. L., Díaz-Noci, J., Meso, K., & Masip, P. (2011). Public sphere 2.0? The democratic qualities of citizen debates in online newspapers. *International Journal of Press/Politics*, 16(4), 463–487. <https://doi.org/10.1177/1940161211415849>
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696–735. <https://doi.org/10.2307/412243>
- Sarewitz, D. (2011). The voice of science: let's agree to disagree. *Nature*, 478(7367), 7. <https://doi.org/10.1038/478007a>
- Savolainen, R. (2001). "Living encyclopedia" or idle talk? Seeking and providing consumer information in an Internet newsgroup. *Library and Information Science Research*, 23, 67–90.
- Savolainen, R. (2011). Requesting and providing information in blogs and internet discussion forums. *Journal of Documentation*, 67(5), 863–886. <https://doi.org/10.1108/00220411111164718>
- Savolainen, R. (2013). Strategies for justifying counter-arguments in Q&A discussion. *Journal of Information Science*, 39(4), 544–556. <https://doi.org/10.1177/0165551513478892>
- Savolainen, R. (2014). The use of rhetorical strategies in Q&A discussion. *Journal of Documentation*, 70(1), 93–118. <https://doi.org/10.1108/JD-11-2012-0152>
- Schegloff, E. A. (1992). Repair After next turn: The last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology*, 97(5), 1295–1345. <https://doi.org/10.1086/229903>

- Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the International Conference on New Methods in Language Processing* (pp. 44–49). <https://doi.org/10.1.1.28.1139>
- Scott, M. (2016). WordSmith Tools. Stroud: Lexical Analysis Software.
- Scott, S. (2002). Linguistic feature variation within disagreements: An empirical investigation. *Text*, 22(2), 301–328. <https://doi.org/10.1515/text.2002.011>
- Secker, J. (2017). The trouble with terminology: rehabilitation and rethinking “digital literacy.” In R. Katharine & J. Parker (Eds.), *Digital Literacy Unpacked* (pp. 3–16). London: Facet Publishing.
- Shanahan, M. (2010). Changing the meaning of peer-to-peer? Exploring online comment spaces as sites of negotiated expertise. *Journal of Science Communication*, 9(1), 1–13.
- Sharples, M., & Ferguson, R. (2019). Pedagogy-informed design of conversational learning at scale. In A. Fessl & T. Z. Draksler (Eds.), *ECTEL Practitioner Proceedings 2019: 14th European Conference on Technology Enhanced Learning* (Vol. 2437).
- Shi, L., & Cristea, A. I. (2018). Demographic indicators influencing learning activities in MOOCs: Learning analytics of FutureLearn courses. In *Proceedings of the 27th International Conference on Information Systems Development (ISD2018), Education Track, Lund, Sweden, August 22-24, 2018*. Cham: Springer. *Lecture notes in information systems and organization*.
- Simon-Vandenberg, A. M. (2000). The functions of I think in political discourse. *International Journal of Applied Linguistics (United Kingdom)*, 10(1), 41–63. <https://doi.org/10.1111/j.1473-4192.2000.tb00139.x>
- Sinclair, J. (2003). *Reading concordances: An Introduction*. London: Pearson.
- Singer, P., Lemmerich, F., West, R., Zia, L., Wulczyn, E., Strohmaier, M., & Leskovec, J. (2017). Why we read Wikipedia. In *2017 International World Wide Web Conference Committee* (pp. 1591–1600).
- Siromaa, M. (2012). Resonance in conversational second stories: A dialogic resource for stance taking. *Text and Talk*, 32(4), 525–545. <https://doi.org/10.1515/text-2012-0025>
- Sotillo, S. M., & Wang-Gempp, J. (2016). Using Corpus Linguistics to Investigate Class, Ideology, and Discursive Practices in Online Political Discussions. *Applied Corpus Linguistics*, 91–122. https://doi.org/10.1163/9789004333772_007
- Springer, N., Engelmann, I., & Pfaffinger, C. (2015). User comments: Motives and inhibitors to write and read. *Information Communication and Society*, 18(7), 798–815. <https://doi.org/10.1080/1369118X.2014.997268>
- Stahl, G. (2015). Conceptualizing the intersubjective group. *International Journal of Computer-Supported Collaborative Learning*, 10(3), 209–217. <https://doi.org/10.4155/fmc-2017-0115>
- Stommel, W., & Koole, T. (2010). The online support group as a community: A micro-analysis of the interaction with a new member. *Discourse Studies*, 12(3), 357–378. <https://doi.org/10.1177/1461445609358518>
- Stubbs, M. (1986). “A matter of prolonged field work”: Notes towards a modal grammar of English. *Applied Linguistics*, 7(1), 1–25. <https://doi.org/10.1093/applin/7.1.1>
- Sudau, F., Friede, T., Grabowski, J., Koschack, J., Makedonski, P., & Himmel, W. (2014). Sources of information and behavioral patterns in online health forums: Observational study. *Journal of Medical Internet Research*, 16(1), e10–e10. <https://doi.org/10.2196/jmir.2875>

- Sunar, A. S., Abdullah, N. A., White, S., & Davis, H. C. (2015). Analysing and predicting recurrent interactions among learners during online discussions in a MOOC. *11th International Conference on Knowledge Management, 4-6 November 2015*.
- Sutherland, J. (2015). Going “meta”: using a metadiscoursal approach to develop secondary students’ dialogic talk in small groups. *Research Papers in Education, 30*(1), 44–69. <https://doi.org/10.1080/02671522.2013.850528>
- Swales, J. M. (2001). Metatalk in American academic talk: The cases of point and thing. *Journal of English Linguistics, 29*(1), 34–54.
- Swinnerton, B., Hotchkiss, S., & Morris, N. P. (2017). Comments in MOOCs: Who is doing the talking and does it help? *Journal of Computer Assisted Learning, 33*(1), 51–64. <https://doi.org/10.1111/jcal.12165>
- Tagg, C. (2012). *The Discourse of Text Messaging: Analysis of Text Message Communication*. London: Bloomsbury.
- Tagg, C., & Seargeant, P. (2019). Context design and critical language/media awareness: Implications for a social digital literacies education. *Linguistics and Education*. <https://doi.org/https://doi.org/10.1016/j.linged.2019.100776>
- Tannen, D. (2007). *Talking Voices: Repetition, dialogue, and imagery in conversational discourse* (2nd ed.). Cambridge University Press.
- Tanskanen, S. (2007). Metapragmatic utterances in computer-mediated interaction. In W. Bublitz & A. Hübler (Eds.), *Metapragmatics in Use* (pp. 87–106). John Benjamins Publishing Company.
- TEI Consortium, eds. (2020). TEI P5: Guidelines for Electronic Text Encoding and Interchange. Retrieved from <https://tei-c.org/guidelines/p5/>
- Thomas, M. J. W. (2002). Learning within incoherent structures: The space of online discussion forums. *Journal of Computer Assisted Learning, 18*(3), 351–366. <https://doi.org/10.1046/j.0266-4909.2002.03800.x>
- Thorne, S. L. (2013). Digital literacies. In R. Margaret & T. Hawkins (Eds.), *Framing Languages and Literacies: Socially Situated Views and Perspectives* (pp. 192–218). Taylor & Francis. <https://doi.org/10.1093/elt/ccr077>
- Tian, Y. (2013). Engagement in online hotel reviews: A comparative study. *Discourse, Context and Media, 2*(4), 184–191. <https://doi.org/10.1016/j.dcm.2013.10.002>
- Tsui, A. B. M. (2009). The pragmatic functions of I don’t know. *Text - Interdisciplinary Journal for the Study of Discourse, 11*(4), 607. <https://doi.org/10.1515/text.1.1991.11.4.607>
- Tubman, P., Oztok, M., & Benachour, P. (2016). Being social or social Learning: A sociocultural analysis of the FutureLearn MOOC platform. In *2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)* (pp. 1–2). IEEE. <https://doi.org/10.1109/ICALT.2016.28>
- Tyrkko, J. (2010). Hyperlinks: Keywords or key words? In M. Bondi & M. Scott (Eds.), *Keyness in Texts* (pp. 79–91). John Benjamins Publishing Company.
- Ugoretz, J. (2005). “Two Roads Diverged in a Wood”: Productive digression in asynchronous discussion. *Innovate - Journal of Online Education, 1*(3), Article 7.
- Vásquez, C. (2018). “Usually not one to complain but...”: constructing identities in user-generated online reviews. In P. Seargeant & C. Tagg (Eds.), *The Language of Social Media: Identity and Community on the Internet* (pp. 65–90). Palgrave Macmillan.

- Veletsianos, G., Kimmons, R., Larsen, R., Dousay, T. A., & Lowenthal, P. R. (2018). Public comment sentiment on educational videos: Understanding the effects of presenter gender, video format, threading, and moderation on YouTube TED talk comments. *Plos One*, 13(6), e0197331. <https://doi.org/10.1371/journal.pone.0197331>
- Vygotsky, L. (1978). *Mind in Society*. Cambridge, MA: Harvard University Press.
- Wagner, D., & Herbel-Eisenmann, B. (2008). “just don’t”: The suppression and invitation of dialogue in the mathematics classroom. *Educational Studies in Mathematics*, 67(2), 143–157. <https://doi.org/10.1007/s10649-007-9097-x>
- Walter, S., Brüggemann, M., & Engesser, S. (2018). Echo chambers of denial: Explaining user comments on climate change. *Environmental Communication*, 12(2), 204–217. <https://doi.org/10.1080/17524032.2017.1394893>
- Wang, X., Wen, M., & Rosé, C. P. (2016). Towards triggering higher-order thinking behaviors in MOOCs. *Proceedings of the 6th International Conference on Analytics and Knowledge*, (November), 398–407. <https://doi.org/http://dx.doi.org/10.1145/2883851.2883964>
- Warchał, K. (2010). Moulding interpersonal relations through conditional clauses: Consensus-building strategies in written academic discourse. *Journal of English for Academic Purposes*, 9(2), 140–150. <https://doi.org/10.1016/j.jeap.2010.02.002>
- Weatherall, A. (2011). I don’t know as a prepositioned epistemic hedge. *Research on Language and Social Interaction*, 44(4), 317–337. <https://doi.org/10.1080/08351813.2011.619310>
- Wegerif, R. (2010). Dialogue and teaching thinking with technology: Opening, expanding and deepening the “inter-face.” In K. Littleton & C. Howe (Eds.), *Educational dialogues: understanding and promoting productive interaction* (pp. 304–322). Routledge.
- Wegerif, R., & Mercer, N. (1997). Using computer-based text analysis to integrate qualitative and quantitative methods in research on collaborative learning. *Language and Education*, 11(4), 271–286. <https://doi.org/10.1080/09500789708666733>
- Wen, M., Yang, D., & Rosé, C. P. (2014). Sentiment analysis in MOOC discussion forums: What does it tell us? In *Educational Data Mining*.
- Wikgren, M. (2001). Health discussions on the Internet : A study of knowledge communication through citations. *Library and Information Science Research*, 23(4), 305–317.
- Wikgren, M. (2003). Everyday health information exchange and citation behaviour in Internet discussion groups. *The New Review of Information Behaviour Research*, 4(1), 225–239. <https://doi.org/10.1080/14716310310001631543>
- Wikipedia. (n.d.). Massive open online course - Wikipedia. Retrieved September 27, 2020, from https://en.wikipedia.org/wiki/Massive_open_online_course
- Wise, A. F., Cui, Y., & Vytasek, J. (2016). Bringing order to chaos in MOOC discussion forums with content-related thread identification. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16* (pp. 188–197). New York, New York, USA: ACM Press. <https://doi.org/10.1145/2883851.2883916>
- Wise, A. F., & Paulus, T. M. (2016). Analyzing learning in online discussions. In C. Haythornthwaite, R. Andrews, J. Fransman, & E. Meyers (Eds.), *The SAGE Handbook of E-learning Research* (2nd ed., pp. 270–290). Sage Publications.
- Wright, S. (2012). From “Third Place” to “Third Space”: Everyday political talk in non-political online spaces. *Javnost - The Public*, 19(3), 5–20.

- Wright, S., & Street, J. (2007). Democracy , deliberation and design : the case of online discussion forums, 9(5), 849–869. <https://doi.org/10.1177/1461444807081230>
- Yang, D., Wen, M., Howley, I., Kraut, R., & Rose, C. (2015). Exploring the effect of confusion in discussion forums of Massive Open Online Courses. *L@S '15 Proceedings of the Second (2015) ACM Conference on Learning @ Scale*, 121–130. <https://doi.org/10.1145/2724660.2724677>
- Yang, W. (2014). Stance and engagement: A corpus-based analysis of academic spoken discourse across science domains. *LSP Journal*, 5(1), 62–78. Retrieved from <http://lsp.cbs.dk>
- Yates, S. J. (1996). Oral and written linguistic aspects of computer conferencing: A Corpus-based study. *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives*, (1991), 29–46. <https://doi.org/10.1075/pbns.39.05yat>
- Zhang, J., Chang, J. P., Danescu-Niculescu-Mizil, C., Dixon, L., Hua, Y., Thain, N., & Taraborelli, D. (2018). Conversations gone awry: Detecting early signs of conversational failure. *Proceedings of ACL 2018*. Retrieved from <http://arxiv.org/abs/1805.05345>
- Ziegele, M., Breiner, T., & Quiring, O. (2014). What creates interactivity in online news discussions? An exploratory analysis of discussion factors in user comments on news items. *Journal of Communication*, 64(6), 1111–1138. <https://doi.org/10.1111/jcom.12123>
- Ziegler, M. F., Paulus, T., & Woodside, M. (2014). Understanding informal group learning in online communities through discourse analysis. *Adult Education Quarterly*, 64(1), 60–78. <https://doi.org/10.1177/0741713613509682>

Appendix

Appendix A Facilitators' comments and wordcount in each MOOC

Abbreviation	Facilitators				
	Number of Facilitators	Comments contributed by facilitators	Average number of comments per facilitator	Wordcount of facilitators' comments	Average wordcount per facilitators' comment
accessibility-2	13	571	43.92	28446	49.82
ancient-1	5	1379	275.80	56555	41.01
code-1	6	1018	169.67	43014	42.25
corpus-1	22	7595	345.23	360866	47.51
dyslexia-1	11	2988	271.64	123806	41.43
finance-1	5	363	72.60	32427	89.33
management-4	6	691	115.17	16373	23.69
moons-1	9	1385	153.89	55966	40.41
nutrition-4	5	260	52.00	10364	39.86
oceans-1	13	1586	122.00	100947	63.65
palliative-1	7	391	55.86	8872	22.69
soils-1	7	809	115.57	43319	53.55

Note. The facilitators' contribution is fewer than those contributed by users, suggesting the discussion space is driven by users. 94 percent of facilitators' contributions are replies, suggesting facilitators might be mainly answering questions. Similarly, the facilitators' contributions vary across the MOOCs, with those MOOCs involving large number of facilitators containing more facilitators' contributions.



From Dr Louise Westmarland
Chair, The Open University Human Research Ethics Committee
Email Louise.westmarland@open.ac.uk
Extension 52462

To Shi Min Chua

Subject Pilot study on the relationship between course contents and learners'

Ref comments
HREC 2016 2368 Chua

AMS (Red)

Submitted 03/11/2016

Date 11/11/2016

Memorandum

This memorandum is to confirm that the research protocol for the above-named research project, as submitted for ethics review, **has been given a favourable opinion** by the Open University Human Research Ethics Committee by **Chair's action** as it is thought to be low risk. Please note that the OU research ethics review procedures are fully compliant with the majority of grant awarding bodies and their Frameworks for Research Ethics.

Please make sure that any question(s) relating to your application and approval are sent to Research-REC-Review@open.ac.uk quoting the HREC reference number above. We will endeavour to respond as quickly as possible so that your research is not delayed in any way.

At the conclusion of your project, by the date that you stated in your application, the Committee would like to receive a summary report on the progress of this project, any ethical issues that have arisen and how they have been dealt with.

Kind regards,

Dr Louise Westmarland

Chair OU HREC

Appendix C Statistics for keyword analysis of initiating posts compared to independent posts

Initiating Keywords	Frequency in Initiating Posts	Frequency in Independent Posts	Expected Frequency in Initiating Posts	Expected Frequency in Independent Posts	Log Likelihood Ratio Test Statistic	p-value	Dispersion Measure	Effect Size
please	589	342	261	670	498.67	1.85 ⁻¹¹⁰	0.21	4.42
anybody	142	96	67	171	103.31	2.86 ⁻²⁴	0.26	3.79
wondering	338	274	172	440	198.20	5.17 ⁻⁴⁵	0.21	3.16
?	14380	12796	7620	19556	7409.92	0 ¹	0.22	2.88
anyone	860	823	472	1211	396.40	3.35 ⁻⁸⁸	0.15	2.68
question	1281	1485	776	1990	415.61	2.20 ⁻⁹²	0.20	2.21
missing	266	317	163	420	81.35	1.89 ⁻¹⁹	0.21	2.15
"	15091	18020	9284	23827	4595.37	0 ¹	0.24	2.15
wonder	990	1281	637	1634	249.85	2.80 ⁻⁵⁶	0.20	1.98
'	14325	18989	9341	23973	3399.26	0 ¹	0.27	1.94
explain	356	483	235	604	79.42	5.03 ⁻¹⁹	0.21	1.89
surely	446	608	296	758	98.21	3.76 ⁻²³	0.29	1.88
sorry	378	528	254	652	77.75	1.17 ⁻¹⁸	0.22	1.84
numbers	289	415	197	507	54.80	1.33 ⁻¹³	0.23	1.79
why	1981	2920	1374	3527	346.41	2.57 ⁻⁷⁷	0.22	1.74
:	5820	8668	4062	10426	984.36	4.51 ⁻²¹⁶	0.29	1.72
says	387	587	273	701	61.61	4.19 ⁻¹⁵	0.18	1.69
does	2825	4307	2000	5132	441.96	4.05 ⁻⁹⁸	0.15	1.68
told	440	671	312	799	68.77	1.10 ⁻¹⁶	0.21	1.68
mean	752	1150	533	1369	116.42	3.84 ⁻²⁷	0.17	1.68
tell	480	743	343	880	71.24	3.17 ⁻¹⁷	0.10	1.66
sort	386	616	281	721	51.24	8.17 ⁻¹³	0.20	1.61
article	907	1472	667	1712	112.83	2.35 ⁻²⁶	0.25	1.58
e.g.	536	875	396	1015	65.15	6.95 ⁻¹⁶	0.26	1.57
came	496	811	366	941	59.90	9.98 ⁻¹⁵	0.19	1.57
(11457	18985	8536	21906	1310.23	6.77 ⁻²⁸⁷	0.18	1.55
called	521	873	391	1003	56.85	4.69 ⁻¹⁴	0.18	1.53
else	569	962	429	1102	59.74	1.08 ⁻¹⁴	0.18	1.52
wrong	589	998	445	1142	61.25	5.03 ⁻¹⁵	0.16	1.51
)	13031	22159	9867	25323	1333.77	5.18 ⁻²⁹²	0.19	1.51
1	1328	2261	1006	2583	135.19	3.00 ⁻³¹	0.28	1.51
tried	698	1212	536	1374	64.95	7.70 ⁻¹⁶	0.28	1.48
cannot	731	1287	566	1452	63.65	1.48 ⁻¹⁵	0.11	1.46
were	4467	7992	3493	8966	358.95	4.76 ⁻⁸⁰	0.24	1.43
perhaps	1060	1981	853	2188	67.06	2.63 ⁻¹⁶	0.15	1.37
whether	881	1663	713	1831	52.50	4.29 ⁻¹³	0.15	1.36
might	1806	3436	1470	3772	102.53	4.26 ⁻²⁴	0.14	1.35
used	3216	6206	2642	6780	166.68	3.93 ⁻³⁸	0.24	1.33
rather	1257	2429	1034	2652	64.56	9.37 ⁻¹⁶	0.09	1.33
-	7348	14257	6058	15547	367.28	7.30 ⁻⁸²	0.05	1.32
example	1323	2574	1093	2804	64.92	7.82 ⁻¹⁶	0.16	1.32

ca	1139	2225	943	2421	54.36	1.67 ⁻¹³	0.13	1.31
any	3212	6369	2686	6895	137.84	7.90 ⁻³²	0.09	1.29
two	1508	3006	1266	3248	62.23	3.05 ⁻¹⁵	0.16	1.29
he	2312	4610	1941	4981	95.21	1.71 ⁻²²	0.26	1.29
;	4620	9221	3881	9960	188.86	5.63 ⁻⁴³	0.20	1.29
say	1476	2983	1250	3209	54.78	1.35 ⁻¹³	0.08	1.27
here	1897	3882	1620	4159	63.57	1.55 ⁻¹⁵	0.12	1.25
seems	1588	3256	1358	3486	52.35	4.65 ⁻¹³	0.12	1.25
if	7404	15301	6366	16339	228.02	1.61 ⁻⁵¹	0.09	1.24
then	3082	6373	2651	6804	94.42	2.55 ⁻²²	0.05	1.24
did	2978	6198	2573	6603	86.08	1.73 ⁻²⁰	0.12	1.23
same	2121	4447	1842	4726	57.25	3.84 ⁻¹⁴	0.11	1.22
...	4372	9184	3801	9755	115.89	5.01 ⁻²⁷	0.11	1.22
than	3822	8073	3335	8560	96.04	1.12 ⁻²²	0.11	1.22
one	5725	12213	5030	12908	130.15	3.80 ⁻³⁰	0.05	1.20
by	6674	14325	5888	15111	142.20	8.80 ⁻³³	0.09	1.20
there	7756	16917	6918	17755	137.82	7.98 ⁻³²	0.10	1.18
n't	8299	18264	7448	19115	132.20	1.35 ⁻³⁰	0.08	1.17
just	3917	8668	3529	9056	58.14	2.45 ⁻¹⁴	0.08	1.16
was	12123	26836	10924	28035	179.16	7.39 ⁻⁴¹	0.14	1.16
could	4447	9857	4011	10293	64.60	9.20 ⁻¹⁶	0.10	1.16
's	8272	18647	7548	19371	94.79	2.12 ⁻²²	0.10	1.14
would	9174	20765	8395	21544	98.77	2.83 ⁻²³	0.10	1.13
or	10290	23340	9429	24201	107.21	4.00 ⁻²⁵	0.06	1.13
the	105000	244174	97904	251270	704.53	3.10 ⁻¹⁵⁵	0.06	1.10
on	14461	34518	13733	35246	53.04	3.26 ⁻¹³	0.09	1.08
that	26521	63640	25280	64881	83.82	5.42 ⁻²⁰	0.03	1.07
,	73311	179123	70780	181654	124.91	5.33 ⁻²⁹	0.05	1.05

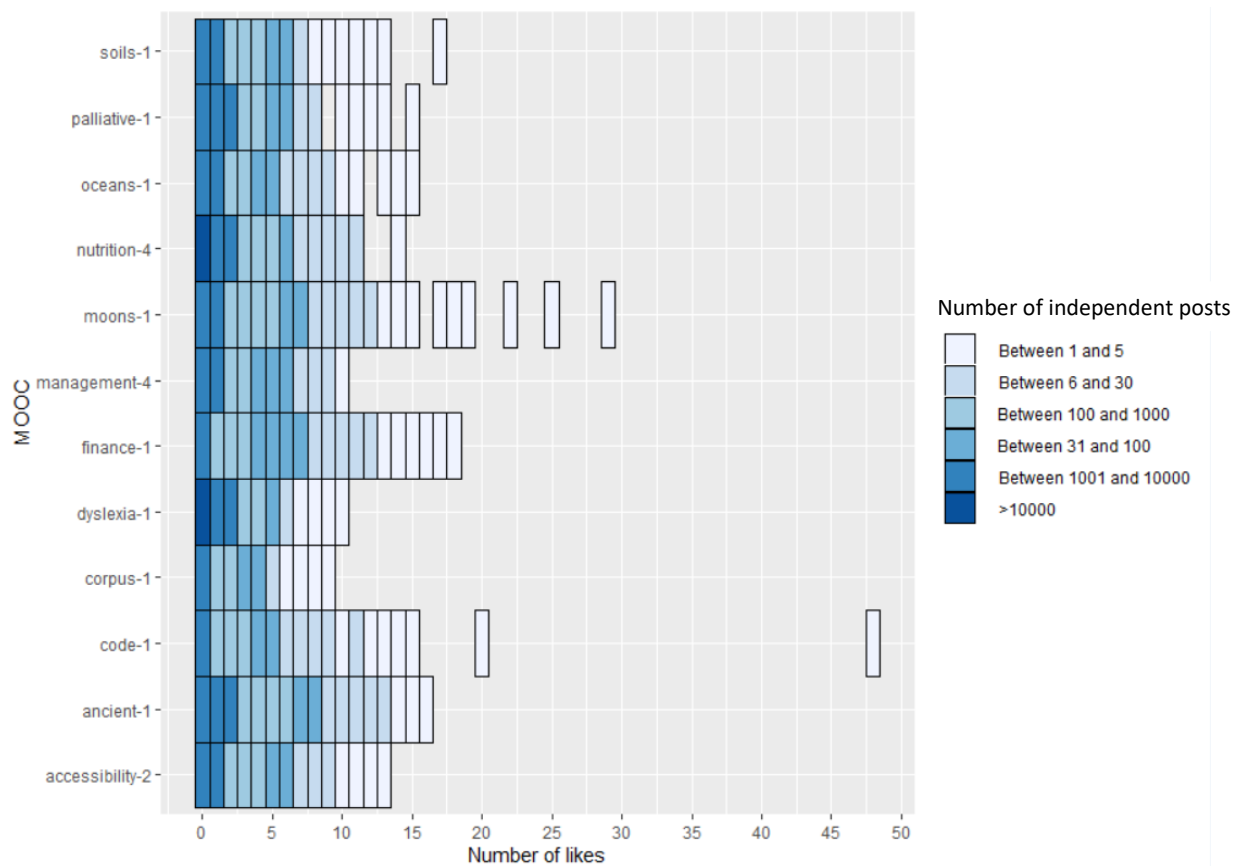
¹ The *p*-value is so small such that the calculation in R only produces 0.

Appendix D Statistics for keyword analysis of independent posts compared to initiating posts

Independent Keywords	Frequency in Initiating Posts	Frequency in Independent Posts	Expected Frequency in Initiating Posts	Expected Frequency in Independent Posts	Log Likelihood Ratio Test Statistic	p-value	Dispersion Measure	Effect Size
joined	79	465	153	391	56.13	6.78 ⁻¹⁴	0.20	2.67
informative	225	1375	449	1151	177.57	1.64 ⁻⁴⁰	0.19	2.44
forward	736	4374	1433	3677	537.41	6.90 ⁻¹¹⁹	0.16	2.29
improve	330	1919	631	1618	226.46	3.53 ⁻⁵¹	0.21	2.21
keen	121	648	216	553	64.76	8.47 ⁻¹⁶	0.22	2.20
hoping	174	872	293	753	74.86	5.05 ⁻¹⁸	0.25	2.00
everyone	872	4364	1468	3768	373.40	3.40 ⁻⁸³	0.20	1.97
knowledge	829	4115	1386	3558	345.10	4.94 ⁻⁷⁷	0.18	1.91
meet	213	1038	351	900	83.11	7.74 ⁻²⁰	0.28	1.89
currently	370	1695	579	1486	114.74	8.98 ⁻²⁷	0.26	1.87
achieve	168	772	264	676	52.72	3.84 ⁻¹³	0.24	1.86
affects	111	579	193	497	54.59	1.48 ⁻¹³	0.29	1.80
opportunity	297	1303	449	1151	77.39	1.40 ⁻¹⁸	0.30	1.75
enjoyed	439	1927	663	1703	114.65	9.40 ⁻²⁷	0.13	1.72
thank	1184	5156	1778	4562	299.07	5.25 ⁻⁶⁷	0.23	1.71
definitely	398	1787	613	1572	114.01	1.30 ⁻²⁶	0.26	1.71
gain	204	942	321	825	65.23	6.65 ⁻¹⁶	0.16	1.67
important	1323	5617	1946	4994	299.56	4.10 ⁻⁶⁷	0.24	1.65
feeling	260	1132	390	1002	65.62	5.47 ⁻¹⁶	0.27	1.64
understanding	653	2704	941	2416	132.10	1.42 ⁻³⁰	0.23	1.63
environment	555	2252	787	2020	102.04	5.43 ⁻²⁴	0.29	1.61
enjoy	291	1196	417	1070	56.85	4.70 ⁻¹⁴	0.25	1.58
helps	326	1365	474	1217	69.38	8.12 ⁻¹⁷	0.22	1.57
looking	1362	5462	1913	4911	236.54	2.24 ⁻⁵³	0.15	1.56
hope	821	3254	1143	2932	134.49	4.27 ⁻³¹	0.23	1.56
yes	486	1882	664	1704	70.61	4.35 ⁻¹⁷	0.25	1.55
great	1716	6781	2382	6115	276.89	3.57 ⁻⁶²	0.18	1.55
aware	510	1952	690	1772	69.60	7.27 ⁻¹⁷	0.22	1.52
good	2953	11543	4065	10431	450.67	5.17 ⁻¹⁰⁰	0.12	1.52
education	589	2314	814	2089	92.26	7.60 ⁻²²	0.29	1.52
excellent	381	1486	523	1344	57.48	3.41 ⁻¹⁴	0.24	1.50
main	437	1645	584	1498	54.38	1.65 ⁻¹³	0.20	1.50
information	1272	4878	1724	4426	175.43	4.82 ⁻⁴⁰	0.20	1.49
mind	757	2901	1026	2632	104.01	2.01 ⁻²⁴	0.24	1.47
thanks	1085	4012	1429	3668	121.79	2.57 ⁻²⁸	0.23	1.44
course	4134	15270	5441	13963	461.11	2.76 ⁻¹⁰²	0.19	1.44
am	5227	19353	6892	17688	591.22	1.36 ⁻¹³⁰	0.17	1.44
learned	554	2007	718	1843	54.92	1.26 ⁻¹³	0.29	1.43
week	1731	6262	2241	5752	170.06	7.16 ⁻³⁹	0.16	1.42
agree	639	2318	829	2128	63.87	1.33 ⁻¹⁵	0.14	1.41
feel	1461	5234	1877	4818	134.93	3.41 ⁻³¹	0.24	1.39
easy	681	2383	859	2205	53.80	2.22 ⁻¹³	0.21	1.39

very	6309	22253	8008	20554	526.06	2.03 ⁻¹¹⁶	0.13	1.37
lot	2065	7267	2617	6715	169.55	9.28 ⁻³⁹	0.15	1.37
'm	3882	13612	4905	12589	311.00	1.32 ⁻⁶⁹	0.23	1.36
every	862	2995	1081	2776	64.81	8.25 ⁻¹⁶	0.12	1.36
love	819	2827	1022	2624	58.77	1.77 ⁻¹⁴	0.18	1.35
my	11940	40692	14757	37875	780.39	9.90 ⁻¹⁷²	0.15	1.33
!	8997	30574	11095	28476	575.45	3.67 ⁻¹²⁷	0.20	1.32
working	1212	4070	1481	3801	70.73	4.10 ⁻¹⁷	0.29	1.32
really	3179	10685	3887	9977	186.84	1.55 ⁻⁴²	0.18	1.31
interesting	3148	10530	3835	9843	178.11	1.26 ⁻⁴⁰	0.10	1.31
better	1439	4806	1751	4494	80.42	3.03 ⁻¹⁹	0.12	1.30
difficult	1273	4105	1508	3870	52.65	3.99 ⁻¹³	0.30	1.26
well	2578	8343	3062	7859	110.16	9.05 ⁻²⁶	0.07	1.26
able	1341	4320	1587	4074	54.96	1.23 ⁻¹³	0.21	1.25
will	4976	15806	5827	14955	178.37	1.10 ⁻⁴⁰	0.10	1.24
always	1656	5257	1938	4975	59.01	1.57 ⁻¹⁴	0.09	1.23
our	3531	11127	4110	10548	116.89	3.04 ⁻²⁷	0.19	1.23
new	1982	6280	2317	5945	69.32	8.39 ⁻¹⁷	0.21	1.23
think	5304	16682	6165	15821	172.16	2.50 ⁻³⁹	0.10	1.23
need	2703	8447	3126	8024	82.05	1.32 ⁻¹⁹	0.19	1.21
and	54344	167252	62133	159463	1394.42	3.42 ⁻³⁰⁵	0.05	1.20
work	3164	9724	3614	9274	79.88	3.97 ⁻¹⁹	0.21	1.20
i	58806	178962	66667	171101	1321.59	2.30 ⁻²⁸⁹	0.16	1.19
more	8198	24880	9275	23803	178.12	1.25 ⁻⁴⁰	0.09	1.18
about	6961	21059	7856	20164	145.38	1.77 ⁻³³	0.09	1.18
also	5260	15800	5905	15155	100.23	1.35 ⁻²³	0.07	1.17
much	3607	10746	4024	10329	61.52	4.38 ⁻¹⁵	0.09	1.16
their	5513	16302	6117	15698	84.57	3.72 ⁻²⁰	0.25	1.15
.	89127	258972	97603	250496	1041.99	1.34 ⁻²²⁸	0.03	1.13
like	4783	13864	5228	13419	53.69	2.34 ⁻¹³	0.09	1.13
with	16384	47576	17934	46026	189.54	4.00 ⁻⁴³	0.14	1.13
all	7305	21055	7952	20408	74.39	6.41 ⁻¹⁸	0.09	1.13
have	18140	50994	19384	49750	112.52	2.75 ⁻²⁶	0.06	1.10
to	65019	182037	69272	177784	367.53	6.47 ⁻⁸²	0.06	1.09
for	20311	56675	21586	55400	105.96	7.52 ⁻²⁵	0.07	1.09

Appendix E Number of likes received by independent posts



Appendix F Statistics for keyword analysis of replies when compared to initiating posts

Words	Frequency in replies	Frequency in Initiating Posts	Expected Frequency in Replies	Expected Frequency in Initiating Posts	Log Likelihood Ratio Test Statistic	p-value	Dispersion Measure	Effect Size
jane	171	11	94	88	157.82	3.40 ⁻³⁶	0.28	14.45
reply	427	45	245	227	329.99	9.68 ⁻⁷⁴	0.29	8.82
michael	237	30	138	129	167.73	2.32 ⁻³⁸	0.26	7.34
agree	4132	639	2473	2298	2607.83	0 ²	0.26	6.01
yes	3029	486	1822	1693	1867.26	0 ²	0.10	5.79
ah	188	31	113	106	113.83	1.42 ⁻²⁶	0.25	5.64
:-	704	158	447	415	335.05	7.62 ⁻⁷⁵	0.28	4.14
agreed	241	57	154	144	109.20	1.47 ⁻²⁵	0.21	3.93
posting	130	31	83	78	58.44	2.10 ⁻¹⁴	0.21	3.90
thanks	4497	1085	2893	2689	1998.40	0 ²	0.19	3.85
hi	2646	692	1730	1608	1082.06	2.62 ⁻²³⁷	0.20	3.55
your	6341	1801	4219	3923	2361.92	0	0.12	3.27
luck	312	101	214	199	98.21	3.76 ⁻²³	0.19	2.87
totally	588	212	415	385	157.51	3.97 ⁻³⁶	0.26	2.58
john	329	120	233	216	86.49	1.41 ⁻²⁰	0.20	2.55
you	22080	8168	15676	14572	5670.73	0	0.11	2.51
comment	874	327	622	579	220.23	8.05 ⁻⁵⁰	0.05	2.48
link	1393	539	1001	931	331.11	5.51 ⁻⁷⁴	0.20	2.40
thank	3038	1184	2188	2034	712.88	4.72 ⁻¹⁵⁷	0.22	2.39
absolutely	377	153	275	255	82.07	1.31 ⁻¹⁹	0.14	2.29
exactly	544	225	399	370	114.16	1.20 ⁻²⁶	0.13	2.25
're	1075	470	801	744	201.27	1.10 ⁻⁴⁵	0.15	2.13
oh	391	172	292	271	72.25	1.90 ⁻¹⁷	0.19	2.11
glad ¹	356	160	267	249	62.74	2.36 ⁻¹⁵	0.12	2.07
mine	339	154	255	238	58.31	2.24 ⁻¹⁴	0.14	2.05
true	870	406	661	615	140.49	2.08 ⁻³²	0.12	1.99
indeed	561	264	428	397	88.78	4.41 ⁻²¹	0.15	1.98
&	1961	957	1512	1406	283.22	1.49 ⁻⁶³	0.26	1.90
post	559	277	433	403	77.55	1.29 ⁻¹⁸	0.10	1.88
too	4330	2151	3359	3122	596.68	8.81 ⁻¹³²	0.11	1.87
'll	976	494	762	708	127.79	1.25 ⁻²⁹	0.19	1.84
point	1929	984	1510	1403	247.10	1.12 ⁻⁵⁵	0.10	1.82
sorry	733	378	576	535	91.02	1.42 ⁻²¹	0.13	1.80
!	16389	8997	13156	12230	1678.27	0 ²	0.12	1.69
maybe	1510	871	1234	1147	130.15	3.80 ⁻³⁰	0.09	1.61
above	775	460	640	595	59.89	1.00 ⁻¹⁴	0.17	1.57
certainly ¹	764	457	633	588	57.25	3.85 ⁻¹⁴	0.11	1.55
right	2076	1275	1737	1614	139.34	3.72 ⁻³²	0.05	1.51
same	3331	2121	2825	2627	189.75	3.61 ⁻⁴³	0.15	1.46
good ¹	4395	2953	3808	3540	189.38	4.35 ⁻⁴³	0.13	1.38
said	1283	868	1115	1036	53.15	3.09 ⁻¹³	0.10	1.37
probably	1240	839	1077	1002	51.34	7.79 ⁻¹³	0.13	1.37

...	6441	4372	5604	5209	261.74	7.15 ⁻⁵⁹	0.15	1.37
's	11909	8272	10459	9722	420.49	1.91 ⁻⁹³	0.08	1.34
just	5535	3917	4898	4554	172.84	1.77 ⁻³⁹	0.06	1.31
say	2075	1476	1840	1711	62.54	2.61 ⁻¹⁵	0.06	1.31
go	1970	1429	1761	1638	51.52	7.09 ⁻¹³	0.05	1.28
think ¹	7198	5304	6479	6023	166.46	4.39 ⁻³⁸	0.06	1.26
n't	11185	8299	10097	9387	244.38	4.37 ⁻⁵⁵	0.05	1.25
those	2511	1868	2269	2110	53.66	2.38 ⁻¹³	0.16	1.25
well ¹	3346	2578	3070	2854	51.68	6.52 ⁻¹³	0.10	1.21
no	4067	3157	3744	3480	58.14	2.44 ⁻¹⁴	0.10	1.20
did	3815	2978	3520	3273	51.35	7.72 ⁻¹³	0.13	1.19
if	9324	7404	8669	8059	103.01	3.34 ⁻²⁴	0.07	1.17
had	5363	4277	4996	4644	56.18	6.61 ⁻¹⁴	0.15	1.17
it	34354	27419	32013	29760	356.33	1.77 ⁻⁷⁹	0.06	1.16
people	5565	4465	5198	4832	53.96	2.05 ⁻¹³	0.26	1.16
-	9158	7348	8554	7952	88.76	4.46 ⁻²¹	0.07	1.16
do	10904	8750	10185	9469	105.51	9.43 ⁻²⁵	0.08	1.16
;	5736	4620	5367	4989	52.84	3.62 ⁻¹³	0.15	1.15
that	32480	26521	30576	28425	246.54	1.48 ⁻⁵⁵	0.04	1.14
but	14686	12411	14043	13054	61.27	4.97 ⁻¹⁵	0.03	1.10
. ¹	103674	89127	99917	92884	293.66	7.93 ⁻⁶⁶	0.02	1.08

¹ The keywords are not keywords of replies when compared to independent posts.

² The *p*-value is so small such that the calculation in R only produces 0.

Appendix G Statistics for keyword analysis of replies when compared to independent posts

Words	Frequency in replies	Frequency in independent posts	Expected frequency in replies	Expected frequency in independent posts	Log Likelihood Ratio Test Statistic	p-value	Dispersion Measure	Effect Size
the ¹	108459	244174	104149	248484	251.14	1.47 ⁻⁵⁶	0.06	1.06
on ¹	15508	34518	14775	35251	51.12	8.67 ⁻¹³	0.06	1.07
are ¹	16920	37532	16082	38370	61.33	4.83 ⁻¹⁵	0.11	1.08
, ¹	81665	179123	77023	183765	392.58	2.27 ⁻⁸⁷	0.03	1.09
they ¹	11288	24178	10475	24991	88.32	5.57 ⁻²¹	0.14	1.11
but	14686	30596	13374	31908	179.43	6.47 ⁻⁴¹	0.03	1.15
not ¹	15086	31376	13722	32740	188.82	5.75 ⁻⁴³	0.06	1.15
would ¹	9999	20765	9086	21678	127.80	1.24 ⁻²⁹	0.11	1.15
who ¹	4316	8928	3912	9332	58.23	2.33 ⁻¹⁴	0.24	1.15
people	5565	11488	5037	12016	77.19	1.55 ⁻¹⁸	0.26	1.16
only ¹	3655	7541	3307	7889	51.07	8.91 ⁻¹³	0.05	1.16
by ¹	7078	14325	6321	15082	125.78	3.43 ⁻²⁹	0.09	1.18
it	34354	69495	30671	73178	613.94	1.56 ⁻¹³⁵	0.06	1.18
where ¹	2916	5880	2598	6198	54.07	1.93 ⁻¹³	0.08	1.18
one ¹	6203	12213	5439	12977	148.46	3.76 ⁻³⁴	0.04	1.21
hi	2646	5199	2317	5528	64.63	9.03 ⁻¹⁶	0.20	1.21
that	32480	63640	28389	67731	815.43	2.38 ⁻¹⁷⁹	0.04	1.22
was ¹	13769	26836	11993	28612	363.68	4.44 ⁻⁸¹	0.14	1.22
even ¹	2929	5610	2522	6017	90.60	1.76 ⁻²¹	0.05	1.25
do	10904	20734	9344	22294	358.76	5.23 ⁻⁸⁰	0.08	1.25
look ¹	1879	3568	1609	3838	62.54	2.61 ⁻¹⁵	0.08	1.26
those	2511	4758	2147	5122	85.06	2.90 ⁻²⁰	0.16	1.26
had	5363	10135	4577	10921	185.69	2.77 ⁻⁴²	0.15	1.26
out ¹	4279	8079	3650	8708	149.28	2.50 ⁻³⁴	0.06	1.26
!	16389	30574	13870	33093	628.67	9.72 ⁻¹³⁹	0.12	1.28
there ¹	9155	16917	7700	18372	377.34	4.72 ⁻⁸⁴	0.08	1.29
system ¹	1213	2217	1013	2417	54.12	1.88 ⁻¹³	0.26	1.31
get ¹	3735	6791	3109	7417	172.84	1.78 ⁻³⁹	0.09	1.31
up ¹	4459	8094	3707	8846	208.71	2.63 ⁻⁴⁷	0.06	1.31
got ¹	1275	2299	1056	2518	62.45	2.74 ⁻¹⁵	0.13	1.32
still ¹	2459	4432	2035	4856	120.79	4.26 ⁻²⁸	0.06	1.32
his ¹	1953	3512	1614	3851	97.40	5.67 ⁻²³	0.26	1.33
though ¹	1632	2933	1348	3217	81.71	1.57 ⁻¹⁹	0.07	1.33
bit ¹	1445	2591	1192	2844	73.45	1.03 ⁻¹⁷	0.11	1.33
no	4067	7266	3347	7986	211.69	5.86 ⁻⁴⁸	0.10	1.34
than ¹	4530	8073	3722	8881	239.63	4.75 ⁻⁵⁴	0.10	1.34
ca ¹	1252	2225	1027	2450	67.42	2.20 ⁻¹⁶	0.06	1.34
problem ¹	1644	2917	1347	3214	89.43	3.18 ⁻²¹	0.14	1.34
go	1970	3443	1599	3814	117.61	2.11 ⁻²⁷	0.05	1.37
were ¹	4614	7992	3723	8883	290.43	4.01 ⁻⁶⁵	0.24	1.38
thank	3038	5156	2420	5774	214.46	1.46 ⁻⁴⁸	0.22	1.41
back ¹	1692	2861	1345	3208	121.86	2.48 ⁻²⁸	0.08	1.41

here ¹	2307	3882	1828	4361	170.52	5.70 ⁻³⁹	0.10	1.42
remember ¹	1130	1885	890	2125	87.41	8.84 ⁻²¹	0.16	1.43
sure ¹	1754	2899	1374	3279	142.21	8.76 ⁻³³	0.06	1.44
book ¹	659	1086	515	1230	54.22	1.79 ⁻¹³	0.24	1.45
least ¹	904	1489	707	1686	74.56	5.88 ⁻¹⁸	0.08	1.45
might	2089	3436	1632	3893	173.51	1.27 ⁻³⁹	0.13	1.45
anything ¹	794	1304	620	1478	66.44	3.60 ⁻¹⁶	0.09	1.45
if	9324	15301	7273	17352	783.24	2.37 ⁻¹⁷²	0.07	1.45
case ¹	1056	1732	823	1965	88.94	4.07 ⁻²¹	0.14	1.45
n't	11185	18264	8698	20751	962.63	2.38 ⁻²¹¹	0.05	1.46
(¹	11682	18985	9057	21610	1028.69	1.04 ⁻²²⁵	0.13	1.47
did	3815	6198	2957	7056	336.44	3.79 ⁻⁷⁵	0.13	1.47
away ¹	803	1304	622	1485	70.97	3.63 ⁻¹⁷	0.19	1.47
then ¹	3927	6373	3042	7258	348.13	1.08 ⁻⁷⁷	0.10	1.47
;	5736	9221	4418	10539	531.64	1.24 ⁻¹¹⁷	0.15	1.48
'd ¹	1095	1760	843	2012	101.57	6.91 ⁻²⁴	0.18	1.48
: ¹	5401	8668	4155	9914	504.46	1.02 ⁻¹¹¹	0.22	1.49
he ¹	2903	4610	2219	5294	284.50	7.84 ⁻⁶⁴	0.24	1.50
thinking ¹	828	1309	631	1506	82.78	9.16 ⁻²⁰	0.12	1.51
actually ¹	1196	1887	911	2172	120.64	4.59 ⁻²⁸	0.09	1.51
rather ¹	1548	2429	1175	2802	159.94	1.17 ⁻³⁶	0.08	1.52
just	5535	8668	4195	10008	576.76	1.90 ⁻¹²⁷	0.06	1.52
's	11909	18647	9025	21531	1241.79	5.06 ⁻²⁷²	0.08	1.52
either ¹	762	1188	576	1374	80.94	2.33 ⁻¹⁹	0.06	1.53
-	9158	14257	6916	16499	978.85	7.10 ⁻²¹⁵	0.07	1.53
let ¹	591	918	446	1063	63.77	1.40 ⁻¹⁵	0.07	1.54
correct ¹	532	822	400	954	58.69	1.84 ⁻¹⁴	0.21	1.54
right	2076	3207	1560	3723	229.24	8.75 ⁻⁵²	0.05	1.54
probably	1240	1915	932	2223	137.09	1.15 ⁻³¹	0.13	1.54
came ¹	533	811	397	947	62.62	2.51 ⁻¹⁵	0.12	1.57
why ¹	1929	2920	1432	3417	231.35	3.03 ⁻⁵²	0.11	1.58
tell ¹	492	743	365	870	59.56	1.19 ⁻¹⁴	0.12	1.58
likely ¹	595	894	440	1049	73.48	1.02 ⁻¹⁷	0.18	1.59
) ¹	14766	22159	10906	26019	1832.20	0 ²	0.14	1.59
completely ¹	471	702	346	827	60.00	9.47 ⁻¹⁵	0.14	1.60
old ¹	1044	1556	768	1832	133.01	8.99 ⁻³¹	0.23	1.60
else ¹	647	962	475	1134	83.19	7.47 ⁻²⁰	0.11	1.60
heard ¹	588	870	431	1027	77.01	1.70 ⁻¹⁸	0.17	1.61
said	1283	1870	931	2222	177.63	1.60 ⁻⁴⁰	0.10	1.64
later ¹	699	1018	507	1210	97.05	6.75 ⁻²³	0.15	1.64
&	1961	2842	1419	3384	277.12	3.19 ⁻⁶²	0.26	1.65
simply ¹	474	685	342	817	67.67	1.94 ⁻¹⁶	0.12	1.65
say	2075	2983	1494	3564	301.77	1.36 ⁻⁶⁷	0.06	1.66
does ¹	3002	4307	2159	5150	439.67	1.28 ⁻⁹⁷	0.07	1.66
off ¹	1146	1635	821	1960	171.14	4.16 ⁻³⁹	0.17	1.67
...	6441	9184	4615	11010	963.85	1.29 ⁻²¹¹	0.15	1.67
surely ¹	429	608	306	731	65.55	5.67 ⁻¹⁶	0.25	1.68
guess ¹	815	1152	581	1386	125.65	3.66 ⁻²⁹	0.15	1.69

went ¹	522	734	371	885	81.91	1.42 ⁻¹⁹	0.18	1.70
wonder ¹	917	1281	649	1549	147.07	7.59 ⁻³⁴	0.14	1.71
anyone ¹	590	823	417	996	95.08	1.83 ⁻²²	0.11	1.71
fine ¹	320	445	226	539	52.09	5.29 ⁻¹³	0.16	1.72
wrong ¹	723	998	508	1213	120.57	4.75 ⁻²⁸	0.07	1.73
perhaps ¹	1442	1981	1011	2412	244.18	4.83 ⁻⁵⁵	0.12	1.74
saw ¹	403	550	281	672	69.68	6.96 ⁻¹⁷	0.15	1.75
ok ¹	441	600	307	734	77.00	1.71 ⁻¹⁸	0.22	1.75
worth ¹	436	592	304	724	76.61	2.08 ⁻¹⁸	0.13	1.76
too	4330	5876	3014	7192	762.13	9.24 ⁻¹⁶⁸	0.11	1.76
nothing ¹	591	800	411	980	104.84	1.32 ⁻²⁴	0.13	1.76
'll	976	1316	677	1615	175.24	5.32 ⁻⁴⁰	0.19	1.77
called ¹	652	873	450	1075	119.59	7.77 ⁻²⁸	0.14	1.78
sort ¹	461	616	318	759	85.08	2.86 ⁻²⁰	0.12	1.79
same	3331	4447	2297	5481	616.43	4.47 ⁻¹³⁶	0.15	1.79
suggest ¹	325	433	224	534	60.52	7.30 ⁻¹⁵	0.08	1.79
running ¹	321	427	221	527	60.05	9.24 ⁻¹⁵	0.24	1.79
numbers ¹	312	415	215	512	58.38	2.16 ⁻¹⁴	0.25	1.79
wo ¹	332	439	228	543	63.23	1.84 ⁻¹⁵	0.14	1.80
round ¹	291	380	198	473	57.48	3.41 ⁻¹⁴	0.21	1.83
suppose ¹	429	557	291	695	86.15	1.66 ⁻²⁰	0.21	1.84
answer ¹	761	987	516	1232	153.30	3.30 ⁻³⁵	0.21	1.84
reference ¹	292	375	197	470	60.48	7.42 ⁻¹⁵	0.28	1.86
explain ¹	377	483	254	606	78.62	7.55 ⁻¹⁹	0.25	1.86
' ¹	14962	18989	10027	23924	3202.34	0 ²	0.18	1.88
mine ¹	339	429	227	541	73.13	1.21 ⁻¹⁷	0.14	1.89
maybe	1510	1909	1010	2409	326.63	5.22 ⁻⁷³	0.09	1.89
friend ¹	281	353	187	447	61.84	3.73 ⁻¹⁵	0.17	1.90
older ¹	439	550	292	697	97.31	5.94 ⁻²³	0.26	1.90
absolutely	377	469	250	596	85.15	2.77 ⁻²⁰	0.14	1.92
suspect ¹	265	327	175	417	61.14	5.31 ⁻¹⁵	0.17	1.93
anyway ¹	334	411	220	525	77.62	1.25 ⁻¹⁸	0.11	1.94
mean ¹	939	1150	617	1472	220.91	5.72 ⁻⁵⁰	0.14	1.95
bottom ¹	252	307	165	394	60.09	9.05 ⁻¹⁵	0.24	1.96
" ¹	14800	18020	9693	23127	3534.33	0 ²	0.19	1.96
told ¹	553	671	362	862	133.22	8.10 ⁻³¹	0.23	1.97
says ¹	484	587	316	755	116.74	3.28 ⁻²⁷	0.18	1.97
mention ¹	359	423	231	551	92.98	5.28 ⁻²²	0.11	2.02
question ¹	1262	1485	811	1936	327.90	2.75 ⁻⁷³	0.21	2.03
presumably ¹	195	229	125	299	50.91	9.67 ⁻¹³	0.25	2.03
happened ¹	264	306	168	402	71.09	3.41 ⁻¹⁷	0.22	2.06
posted ¹	191	211	119	283	57.32	3.71 ⁻¹⁴	0.20	2.16
Indeed	561	608	345	824	175.37	4.97 ⁻⁴⁰	0.15	2.20
? ¹	11811	12796	7268	17339	3694.87	0 ²	0.09	2.20
above	775	836	476	1135	244.66	3.79 ⁻⁵⁵	0.17	2.21
point	1929	2061	1178	2812	621.19	4.11 ⁻¹³⁷	0.10	2.23
statement ¹	184	195	112	267	60.25	8.36 ⁻¹⁵	0.22	2.25
saying ¹	528	557	320	765	174.50	7.68 ⁻⁴⁰	0.11	2.26

original ¹	292	304	176	420	99.09	2.41 ⁻²³	0.25	2.29
apparently ¹	249	257	149	357	85.95	1.85 ⁻²⁰	0.25	2.31
totally	588	606	353	841	203.55	3.51 ⁻⁴⁶	0.26	2.31
somewhere ¹	265	271	158	378	93.13	4.89 ⁻²²	0.14	2.33
worry ¹	204	206	121	289	73.45	1.03 ⁻¹⁷	0.13	2.36
meant ¹	387	390	229	548	139.87	2.84 ⁻³²	0.13	2.37
sent ¹	146	147	87	206	52.86	3.58 ⁻¹³	0.15	2.37
opposite ¹	171	172	101	242	62.03	3.39 ⁻¹⁵	0.24	2.37
suggestion ¹	163	163	96	230	59.78	1.06 ⁻¹⁴	0.18	2.39
re ¹	233	233	138	328	85.45	2.38 ⁻²⁰	0.18	2.39
luck	312	310	184	438	115.80	5.27 ⁻²⁷	0.19	2.40
response ¹	248	242	145	345	95.14	1.78 ⁻²²	0.11	2.45
post	559	519	318	760	233.93	8.29 ⁻⁵³	0.10	2.57
posts ¹	176	161	100	237	75.50	3.65 ⁻¹⁸	0.21	2.61
true	870	792	491	1171	376.23	8.23 ⁻⁸⁴	0.12	2.62
you	22080	19849	12384	29545	9746.85	0 ²	0.11	2.65
thanks	4497	4012	2513	5996	2009.60	0 ²	0.19	2.67
wondering ¹	308	274	172	410	138.27	6.37 ⁻³²	0.17	2.68
press ¹	139	120	76	183	65.41	6.09 ⁻¹⁶	0.28	2.76
're	1075	919	589	1405	513.49	1.11 ⁻¹¹³	0.15	2.79
agreed	241	195	129	307	124.81	5.61 ⁻²⁹	0.21	2.95
mark ¹	223	176	118	281	119.58	7.82 ⁻²⁸	0.23	3.02
exactly	544	422	285	681	298.66	6.45 ⁻⁶⁷	0.13	3.08
oh	391	291	201	481	226.72	3.09 ⁻⁵¹	0.19	3.21
sorry	733	528	372	889	442.95	2.47 ⁻⁹⁸	0.13	3.31
your	6341	4506	3204	7643	3896.47	0 ²	0.12	3.36
wondered ¹	194	137	98	233	120.12	5.95 ⁻²⁸	0.25	3.38
please ¹	502	342	249	595	324.40	1.59 ⁻⁷²	0.14	3.50
link	1393	941	689	1645	909.09	1.04 ⁻¹⁹⁹	0.20	3.53
comment	874	575	428	1021	587.83	7.43 ⁻¹³⁰	0.05	3.63
yes	3029	1882	1450	3461	2168.24	0 ²	0.10	3.84
:- ¹	704	408	328	784	541.03	1.12 ⁻¹¹⁹	0.28	4.12
john	329	187	152	364	257.68	5.51 ⁻⁵⁸	0.20	4.20
agree	4132	2318	1905	4545	3277.16	0 ²	0.26	4.25
paul ¹	182	91	81	192	160.11	1.07 ⁻³⁶	0.28	4.77
posting	130	51	53	128	137.55	9.13 ⁻³²	0.21	6.08
ah	188	59	73	174	228.29	1.41 ⁻⁵¹	0.25	7.60
michael	237	64	89	212	311.42	1.07 ⁻⁶⁹	0.26	8.84
jane	171	28	59	140	275.02	9.12 ⁻⁶²	0.28	14.57
reply	427	48	140	335	764.13	3.40 ⁻¹⁶⁸	0.29	21.22

¹ The keywords marked with 1 are not keywords of replies when compared to initiating posts.

² The *p*-value is so small such that the calculation in R only produces 0.

Appendix H Statistics for keyword analysis of reply keywords comparing first contributions and subsequent contributions

Reply Keywords	Frequency in first contributions	Frequency in subsequent contributions	Expected frequency in first contributions	Expected frequency in subsequent contributions	Log Likelihood Ratio Test Statistic	p-value	Effect Size
<u>Reply keywords used more frequently in the first-time contributions</u>							
agree	3430	702	2743	1389	574.75	5.19 ⁻¹²⁷	2.47
too	3206	1124	2875	1455	118.67	1.24 ⁻²⁷	1.44
same	2500	831	2211	1120	117.82	1.90 ⁻²⁷	1.52
&	1513	448	1302	659	108.76	1.83 ⁻²⁵	1.71
<u>Reply keywords used more frequently in the subsequent contributions</u>							
reply	95	332	283	144	349.19	6.36 ⁻⁷⁸	6.90
thank	1033	2005	2017	1021	1323.63	8.29 ⁻²⁹⁰	3.83
ah	67	121	125	63	73.87	8.36 ⁻¹⁸	3.57
thanks	1606	2891	2986	1511	1758.41	0 ¹	3.56
sorry	278	455	487	246	247.00	1.17 ⁻⁵⁵	3.23
oh	181	210	260	131	66.35	3.78 ⁻¹⁶	2.29
'll	484	492	648	328	116.47	3.74 ⁻²⁷	2.01
:-	375	329	467	237	51.73	6.36 ⁻¹³	1.73
yes	1818	1211	2011	1018	53.61	2.44 ⁻¹³	1.32
...	3974	2467	4276	2165	62.25	3.03 ⁻¹⁵	1.23
but	9249	5437	9750	4936	75.47	3.72 ⁻¹⁸	1.16
you	14137	7943	14659	7421	54.73	1.38 ⁻¹³	1.11

¹ The p-value is so small such that the calculation in R only produces 0.

Appendix I Statistics for keyword analysis of reply keywords comparing short threads and long threads.

Reply Keywords	Frequency in short threads	Frequency in long threads	Expected frequency in short threads	Expected frequency in long threads	Log Likelihood Ratio Test Statistic	p-value	Effect Size
agree	3057	1075	2445	1687	396.96	2.53 ⁻⁸⁸	1.96
!	10720	5669	9698	6691	268.98	1.89 ⁻⁶⁰	1.30
yes	2073	956	1792	1237	110.98	5.97 ⁻²⁶	1.50
too	2887	1443	2562	1768	103.28	2.91 ⁻²⁴	1.38
you	13720	8360	13065	9015	81.05	2.20 ⁻¹⁹	1.13
hi	1773	873	1566	1080	68.92	1.03 ⁻¹⁶	1.40
same	2197	1134	1971	1360	64.79	8.31 ⁻¹⁶	1.34

Note. Seven reply keywords are used significantly more often in short threads, while none are found used significantly more often in long threads.

Appendix J Statistics for keyword analysis of reply keywords comparing the first reply of one-reply threads and that of threads with more than one reply.

Reply Keywords	Frequency in first reply of threads	Frequency in first reply of threads with more than one reply	Expected frequency in first reply of threads	Expected frequency in first reply of threads with more than one reply	Log Likelihood Ratio Test Statistic	p-value	Effect Size
!	2910	2934	3316	2528	100.61	1.12 ⁻²³	1.30
&	420	316	418	318	56.41	5.87 ⁻¹⁴	1.74
agree	1275	1013	1298	990	142.98	5.93 ⁻³³	1.65
yes	621	509	641	489	62.17	3.15 ⁻¹⁵	1.60

Note. Four reply keywords are used significantly more often in the first reply of one-reply threads, while none are used significantly more often in the first reply of other threads.